

Least-Squares Filter versus Hodrick-Prescott Filter

by

HANS-JÜRIG BÜTTLER *
Research, Swiss National Bank #, Switzerland
and
Department of Economics, University of Zurich, Switzerland

ABSTRACT

This paper introduces a new filter, called least-squares (LS) filter, as an alternative to the widely used Hodrick-Prescott (HP) filter. The LS filter has two properties which distinguish it from the HP filter. First, the LS filter rather than the HP filter uses first differences as a measure of the smoothness of a nowhere differentiable random variable. Second, the LS filter rather than the HP filter uses a reference time series to avoid the subjective choice of weights.

First version: September 2005
Revised version: November 2007

appeared in *Economic and Financial Modelling*, 2007, 14 (4), 153 – 183.

KEYWORDS: Least-squares filter, Hodrick-Prescott filter

* I would like to thank Karim Abadir and Katrin Assenmacher for valuable comments. The computer programs in Matlab code can be obtained from the author upon request.

Mailing address: Swiss National Bank, P. O. Box, 8022 Zurich, Switzerland.

Phone (direct dialling): +41-1-631 34 17, Fax: +41-1-631 39 01, E-Mail: hans-juerg.buettler@snb.ch.

0 Introduction

This paper introduces the least-squares (LS) filter as an alternative to the widely used Hodrick-Prescott (HP) filter. The LS filter has two properties which distinguish it from the HP filter. First, the LS filter rather than the HP filter uses first differences as a measure of the smoothness of a nowhere differentiable random variable. Second, the LS filter rather than the HP filter uses a reference time series to avoid the subjective choice of weights.

To motivate the introduction of the LS filter, consider the following two stochastic differential equations proposed by Cox, Ingersoll and Ross [1985] for their term structure model.

$$\begin{aligned}\pi(t) dt &\equiv \frac{dp(t)}{p(t)} = y(t) dt + \sigma_p \sqrt{y(t)} dz_1(t), \quad 0 < y < \infty, 0 < \sigma_p < 1, \\ dy(t) &= \kappa [\theta - y(t)] dt + \sigma \sqrt{y(t)} dz_2(t), \quad 0 < \kappa, \theta, \sigma < \infty,\end{aligned}\tag{0-1}$$

where π denotes the instantaneous inflation rate, p the consumer price index, y the expected instantaneous inflation rate, σ_p a volatility parameter, z_1 a Gauss-Wiener process, κ the speed of adjustment, θ the long-run ‘equilibrium’ value of the expected inflation rate, σ another volatility parameter and z_2 another Gauss-Wiener process. The two Gauss-Wiener processes may be correlated. The first equation postulates that the instantaneous inflation rate can be decomposed into a stochastic trend rate, given by the expected instantaneous inflation rate, and a stochastic term, which is described by a square-root process. The process of the stochastic trend rate is described by the second equation above. It states that the randomly moving expected instantaneous inflation rate is elastically pulled toward a long-term equilibrium value, θ . Note that the square-root processes of equation (0-1) require that the expected instantaneous inflation rate, y , remains positive.¹

In general, the expected instantaneous inflation rate, y , is an *unobservable* variable. Hence, we would like to estimate a filter of the instantaneous inflation rate, π , which serves as a proxy for the expected instantaneous inflation rate, y . We require that the filter has two properties. First, it should be a contraction of the original time series (π), that is, the filter should not exhibit a phase shift with respect to the original time series. Second, the variation of the filter should not be greater than the variation of a *reference* time series.² What is a reasonable reference time series? For instance, if we apply the LS filter to a bond option, whose term to maturity is about a year, then it seems reasonable to use the one-year inflation rate as a reference. On the other hand, if we apply the LS filter to a model of the term structure of interest rates, whose maximum term to maturity is thirty years, then the thirty-year inflation rate might be a reasonable reference.

¹ The stochastic differential equations (0-1) may be estimated in a forward-looking manner, if there exists a sufficient number of derivative instruments written on bonds and, if start values for p and y are obtainable.

² In other words, the LS filter should look like the original time series, but should have a smaller variation.

The plan of the paper is the following. In the next section, we discuss the unconstrained LS filter which is in the spirit of the HP filter. The unconstrained LS filter serves as a start value for the constrained LS filter optimization, which is discussed in the second section. The constrained LS filter also considers bounds on the filter values. For instance, the expected instantaneous inflation rate, y , in equation (0-1) must be positive everywhere. In order to reduce the instability of the boundary values, optimized weights are introduced in the third section. In the fourth section, the LS filters are compared with the HP filter. The results are illustrated by an example for the observed one-month inflation rate over a sample period of nine years when the reference time series –considering a hypothetical forecasting horizon of five years– is chosen to be the observed five-year inflation rate (see Figure 1).

1 The unconstrained LS filter

Let \mathbf{y} denote the $n \times 1$ vector of observed values of the original time series for which we wish to calculate a LS filter. Let \mathbf{x} denote the $n \times 1$ vector of LS filter values, then the objective function F_1 measures the squared deviations of the LS filter from the original times series as given in equation (1-1).

$$F_1(\mathbf{x}) \equiv (\mathbf{y} - \mathbf{x})' \mathbf{I} (\mathbf{y} - \mathbf{x}) = \sum_{i=1}^n (y_i - x_i)^2, \quad \frac{dF_1}{d\mathbf{x}} = -2 \mathbf{I} (\mathbf{y} - \mathbf{x}) \quad (1-1)$$

where \mathbf{I} is a $n \times n$ identity matrix. Let \mathbf{E} denote the $n \times n$ matrix to obtain the first differences of the LS filter values

$$\mathbf{E} = \begin{bmatrix} -1 & +1 & 0 & \cdots & 0 \\ 0 & -1 & +1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1 & +1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1-2)$$

and let \mathbf{G} denote the $n \times n$ symmetric and positive definite matrix as follows

$$\mathbf{G} = \mathbf{E}' \mathbf{E}, \quad (1-3)$$

then the second objective function F_2 measures the smoothness of the LS filter by squared first differences

$$F_2(\mathbf{x}) \equiv \mathbf{x}' \mathbf{G} \mathbf{x} = \sum_{i=2}^n (x_i - x_{i-1})^2, \quad \frac{dF_2}{d\mathbf{x}} = 2 \mathbf{G} \mathbf{x}. \quad (1-4)$$

Since a Gauss-Wiener process is nowhere differentiable, first differences rather than second differences as used in the HP filter are appropriate. For the same reason, the use of a numeric

approximation to the first derivative or second derivative, respectively, evaluated at the data points would be inappropriate.

The objective of the unconstrained LS filter is to minimize a linear combination of the two functions F_1 and F_2 , denoted as F , with respect to the LS filter values, that is,

$$\min_{\{\mathbf{x}\}} \{F(\mathbf{x}) \equiv w_0 F_1(\mathbf{x}) + (1 - w_0) F_2(\mathbf{x})\}, \quad 0 \leq w_0 \leq 1, \quad (1-5)$$

where w_0 denotes the weight of the linear combination. The linear combination above defines two extreme cases. On the one hand, the squared deviations from the original time series are minimized only, if $w_0 = 1$. In this case, the minimum is least, if $F_1 = 0$, that is, if the LS filter is equal to the original time series. On the other hand, the squared first differences of the LS filter are minimized only, if $w_0 = 0$. In this case, the minimum is least, if $F_2 = 0$, that is, if the LS filter is equal to a zero horizontal line, i. e., it has no variation at all. All other solutions are obtained as a linear combination of these two extreme cases from setting $0 < w_0 < 1$. The optimum solution, denoted as \mathbf{x}^* , is given by the first-order conditions, $dF/d\mathbf{x} = \mathbf{0}$, that is,

$$\mathbf{x}^* = w_0 [w_0 \mathbf{I} + (1 - w_0) \mathbf{G}]^{-1} \mathbf{y} \quad (1-6)$$

The optimum is a minimum, since the Hessian matrix is positive definite by

$$\mathbf{x}' \frac{\partial^2 F}{\partial \mathbf{x}^2} \mathbf{x} = 2 w_0 \mathbf{x}' \mathbf{I} \mathbf{x} + 2 (1 - w_0) \mathbf{x}' \mathbf{G} \mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}. \quad (1-7)$$

As explained above, the optimum solution is indeed $\mathbf{x}^* = \mathbf{y}$, if $w_0 = 1$, and $\mathbf{x}^* = \mathbf{0}$, if $w_0 = 0$. The solution in equation (1-6) for $w_0 = 0.1$ serves as start values of the constrained LS filter.

2 The unweighted constrained LS filter

The unconstrained LS filter depends on the personal choice of the weight of the linear combination. In order to make this choice less subjective, we introduce a reference time series, denoted by the $n \times 1$ vector \mathbf{r} . The objective is to minimize the squared deviations of the LS filter from the original time series subject to the condition that the LS filter should be as smooth as the reference time series at least. As explained in the introductory example, the LS filter may be subject to lower bounds, denoted by the $n \times 1$ vector \mathbf{x}_l , as well as upper bounds, denoted by the $n \times 1$ vector \mathbf{x}_u . Hence, the optimization can be stated in the following way.

$$\begin{aligned}
& \min_{\{\mathbf{x}\}} \{F_1(\mathbf{x}) \equiv (\mathbf{y} - \mathbf{x})' \mathbf{I} (\mathbf{y} - \mathbf{x})\} \\
& \text{subject to} \\
& \quad \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \\
& \quad F_2(\mathbf{x}) \equiv \mathbf{x}' \mathbf{G} \mathbf{x} \leq \mathbf{r}' \mathbf{G} \mathbf{r}
\end{aligned} \tag{2-1}$$

Since a closed-form solution of equation (2-1) does not exist, we use an optimization algorithm³, given the start values of equation (1-6) and the gradients of the two functions F_1 and F_2 in equations (1-1) and (1-4).⁴

In the absence of lower and upper bounds, the result for the example considered is shown in Figure 2 for four overlapping samples of equal size. The first sample comprises the period from April 1999 to October 2007, the second sample comprises the period from March 1999 to September 2007, the third sample comprises the period from February 1999 to August 2007 and the fourth sample comprises the period from January 1999 to July 2007. The four LS filters are quite close to each other in the middle of the sample period under consideration, but diverge considerably at both boundaries of the sample period.

3 The optimized constrained LS filter

In order to reduce the instability mentioned above, we introduce weights for the objective function F_1 in equation (2-1) and proceed in two steps. In the first step, the weights are optimized with respect to two overlapping subsamples. In the second step, the LS filter is calculated given the optimized weights.

In the first step, the following procedure is considered. Let \mathbf{w} denote the $n \times 1$ vector of weights, let \mathbf{W} denote the $n \times n$ diagonal matrix of the weights, i. e., $\mathbf{W} = \text{diag}(\mathbf{w})$ and let \mathbf{u} denote the $n \times 1$ unit vector, i. e., $\mathbf{I} = \text{diag}(\mathbf{u})$. The objective is to minimize the squared deviation of the weights from the unit vector, that is, the deviation from the unweighted constrained LS filter.

Now, divide the sample into two overlapping subsamples. The first subsample comprises the first $n - 1$ data of the sample, whereas the second subsample comprises the last $n - 1$ data of the sample. If the observations of the sample considered are denoted by the indices $\{1, 2, 3, \dots, n - 1, n\}$, then the first subsample comprises the observations given by the indices $\{1, 2, 3, \dots, n - 1\}$ and the second subsample comprises the observations given by the indices $\{2, 3, \dots, n - 1, n\}$. The two subsamples have $n - 2$ observations in common given by the indices $\{2, 3, \dots, n - 1\}$. Let \mathbf{z}_1^* denote the $(n - 1) \times 1$ vector of the optimum LS filter obtained for the first subsample and let \mathbf{z}_2^* denote the $(n - 1) \times 1$ vector of the optimum LS filter ob-

³ We use Matlab's optimization procedure «constr.m». Considering the Kuhn-Tucker conditions, the function «constr.m» applies a sequential quadratic programming (SQP) method.

⁴ The use of analytical gradients in the optimization procedure reduces considerably the computation time.

tained for the second subsample. In order to compare the common values of the two subsamples, the first element of \mathbf{z}_1^* is deleted, whereas the last element of \mathbf{z}_2^* is deleted. The resulting $(n-2) \times 1$ vectors are denoted by $\boldsymbol{\zeta}_1^*$ and $\boldsymbol{\zeta}_2^*$, respectively. The minimization is subject to the condition that the common values of both subsamples do not deviate from each other by more than a given tolerance, ϵ . Hence, the optimization of the first step can be stated as follows.

$$\begin{aligned} & \min_{\{\mathbf{w}\}} \{F_3(\mathbf{w}) \equiv (\mathbf{w} - \mathbf{u})' \mathbf{I} (\mathbf{w} - \mathbf{u})\} \\ & \text{subject to} \\ & -\epsilon \leq \left(F_4(\mathbf{w}) \equiv \frac{\boldsymbol{\zeta}_1^*(\mathbf{w})}{\boldsymbol{\zeta}_2^*(\mathbf{w})} - 1 \right) 100 \leq \epsilon, \text{ where} \\ & \boldsymbol{\zeta}_1^*(\mathbf{w}) \equiv [z_{1,2}^*, z_{1,3}^*, \dots, z_{1,n-1}^*]', \\ & \boldsymbol{\zeta}_2^*(\mathbf{w}) \equiv [z_{2,1}^*, z_{2,2}^*, \dots, z_{2,n-2}^*]'. \end{aligned} \quad (3-1)$$

Since a closed-form solution of equation (3-1) does not exist, we use an optimization algorithm, given the $n \times 1$ gradient of the function F_3 as given below

$$\frac{dF_3}{d\mathbf{w}} = 2 \mathbf{I} (\mathbf{w} - \mathbf{u}) \quad (3-2)$$

and the $n \times (n-2)$ gradient of the function F_4 as below.

$$\frac{dF_4}{d\mathbf{w}} = \frac{100}{\boldsymbol{\zeta}_2^*} \frac{d\boldsymbol{\zeta}_1^*}{d\mathbf{w}} - \frac{100 \boldsymbol{\zeta}_1^*}{(\boldsymbol{\zeta}_2^*)^2} \frac{d\boldsymbol{\zeta}_2^*}{d\mathbf{w}} \quad (3-3)$$

Note that the products on the right-hand side of the above equation are outer products. The gradients on the right-hand side will be derived later.

If the start values of the weights are not specified by the user, we use the inverted weights of the Chebyshev polynomial of the first kind as follows (Abramowitz and Stegun, 1972). Let \mathbf{T} denote the $n \times 1$ vector of the ordinal number of the sample observations, which are mapped linearly onto the interval $(-1 + \eta, 1 - \eta)$ by

$$\begin{aligned} \boldsymbol{\tau} &= a + b \mathbf{T}, \text{ where} \\ b &= 2 \frac{(1 - \eta)}{T_n - T_1}, \quad a = (1 - \eta) - b T_n. \end{aligned} \quad (3-4)$$

The parameter $1 - \eta$ defines the largest weight at the boundaries on the interval $(-1 + \eta, 1 - \eta)$. Let R denote the ratio of the largest weight to the smallest weight, where the smallest weight is set equal to one, then the parameter $(1 - \eta)$ becomes the following expression.

$$(1 - \eta) = \sqrt{\frac{R^2 - 1}{R^2}} \quad (3-5)$$

The symmetric Chebyshev weights, reflected at the time axis and shifted up by the value of the first element plus one, $\mathbf{w}_0 = [w_{0,1}, w_{0,2}, \dots, w_{0,n}]'$, are defined by

$$w_{0,i} = -\frac{1}{\sqrt{(1-\tau_i)^2}} + \frac{1}{\sqrt{(1-\tau_1)^2}} + 1, \quad i = 1, 2, \dots, n. \quad (3-6)$$

The $n \times 1$ vector \mathbf{w}_0 is used as the start values for the optimization of equation (3-1), given a ratio of $R = 10$.

It remains to determine the optimum LS filters of the two overlapping subsamples considered in equation (3-1) for *every* step of the optimization procedure. For each subsample, we solve the simplified optimization of equation (2-1) for a given vector of weights, that is, for $i = 1, 2$ we solve

$$\begin{aligned} & \min_{\{\mathbf{z}_i\}} \{F_{1,i}(\mathbf{z}_i) \equiv (\mathbf{y}_i - \mathbf{z}_i)' \mathbf{W}_i (\mathbf{y}_i - \mathbf{z}_i)\} \\ & \text{subject to} \\ & F_{2,i}(\mathbf{z}_i) \equiv \mathbf{z}_i' \mathbf{G}_i \mathbf{z}_i = \mathbf{r}_i' \mathbf{G}_i \mathbf{r}_i \end{aligned} \quad (3-7)$$

where \mathbf{y}_i is the $(n-1) \times 1$ vector of the original time series of the i th subsample, $\mathbf{W}_i = \text{diag}(\mathbf{w}_i)$ is the $(n-1) \times (n-1)$ diagonal matrix of weights of the i th subsample, \mathbf{G}_i is the $(n-1) \times (n-1)$ symmetric and positive definite matrix of first differences of the i th subsample and \mathbf{r}_i is the $(n-1) \times 1$ vector of the reference time series of the i th subsample. The Lagrangian equation for equation (3-7), L_i , is given by

$$L_i \equiv (\mathbf{y}_i - \mathbf{z}_i)' \mathbf{W}_i (\mathbf{y}_i - \mathbf{z}_i) + \gamma_i [\mathbf{r}_i' \mathbf{G}_i \mathbf{r}_i - \mathbf{z}_i' \mathbf{G}_i \mathbf{z}_i] \quad (3-8)$$

where γ_i is the scalar Lagrange multiplier of the i th subsample. The first-order conditions of equation (3-8) become

$$\begin{bmatrix} \partial L_i / \partial \mathbf{z}_i \\ \partial L_i / \partial \gamma_i \end{bmatrix} = \begin{bmatrix} -2 \mathbf{W}_i (\mathbf{y}_i - \mathbf{z}_i) - 2 \gamma_i \mathbf{G}_i \mathbf{z}_i \\ \mathbf{r}_i' \mathbf{G}_i \mathbf{r}_i - \mathbf{z}_i' \mathbf{G}_i \mathbf{z}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix} \quad (3-9)$$

Since a closed-form solution of equation (3-9) does not exist, we solve for the optimum LS filters of both subsamples, \mathbf{z}_1^* and \mathbf{z}_2^* , by a numerical algorithm to solve a nonlinear equation system, given the $n \times n$ matrix of gradients in equation (3-10).

$$\frac{d^2 L_i}{d[(\mathbf{z}_i', \gamma_i)']^2} = \begin{bmatrix} 2 \mathbf{W}_i' - 2 \gamma_i \mathbf{G}_i' & -2 \mathbf{G}_i \mathbf{z}_i \\ -2 \mathbf{z}_i' \mathbf{G}_i' & 0 \end{bmatrix} \quad (3-10)$$

It remains to determine the gradients on the right-hand side of equation (3-3). From a comparative-static analysis of equation (3-9), it follows that

$$\begin{bmatrix} 0 & -\mathbf{z}_i^{*'} \mathbf{G}_i' \\ -\mathbf{G}_i \mathbf{z}_i^* & \mathbf{W}_i - \gamma_i \mathbf{G}_i \end{bmatrix} \begin{bmatrix} (\partial \gamma_i / \partial \mathbf{w}_i)' \\ (\partial \mathbf{z}_i^* / \partial \mathbf{w}_i)' \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_i \end{bmatrix} \quad (3-11)$$

where \mathbf{D}_i is defined as the $(n-1) \times (n-1)$ diagonal matrix of the deviation of the optimum LS filter from the original time series for the i th subsample, that is, $\mathbf{D}_i = \text{diag}(\mathbf{y}_i - \mathbf{z}_i^*)$. Partitioned inversion yields

$$\text{if } \mathbf{A}_i = \begin{bmatrix} 0 & -\mathbf{z}_i^{*'} \mathbf{G}_i' \\ -\mathbf{G}_i \mathbf{z}_i^* & \mathbf{W}_i - \gamma_i \mathbf{G}_i \end{bmatrix}, \quad \mathbf{B}_i = [\mathbf{W}_i - \gamma_i \mathbf{G}_i]^{-1}, \quad \mu_i = \frac{1}{\mathbf{z}_i^{*'} \mathbf{G}_i' \mathbf{B}_i \mathbf{G}_i \mathbf{z}_i^*}$$

then

$$\mathbf{A}_i^{-1} = \begin{bmatrix} \mu_i & \mu_i \mathbf{z}_i^{*'} \mathbf{G}_i \mathbf{B}_i \\ \mu_i \mathbf{B}_i \mathbf{G}_i \mathbf{z}_i^* & \mathbf{B}_i + \mu_i \mathbf{B}_i \mathbf{G}_i \mathbf{z}_i^* \mathbf{z}_i^{*'} \mathbf{G}_i' \mathbf{B}_i \end{bmatrix} \quad (3-12)$$

From equations (3-11) and (3-12), it follows that the gradient of the optimum LS filter with respect to the weights of the i th subsample is given by

$$\frac{\partial \mathbf{z}_i^*}{\partial \mathbf{w}_i} = ([\mathbf{W}_i - \gamma_i \mathbf{G}_i]^{-1} \mathbf{D}_i + \mu_i [\mathbf{W}_i - \gamma_i \mathbf{G}_i]^{-1} \mathbf{G}_i \mathbf{z}_i^* \mathbf{z}_i^{*'} \mathbf{G}_i' [\mathbf{W}_i - \gamma_i \mathbf{G}_i]^{-1} \mathbf{D}_i)' \quad (3-13)$$

The gradient $d\zeta_1^*/d\mathbf{w}$ is obtained from the gradient $d\mathbf{z}_1^*/d\mathbf{w}_1$ by extending the latter by a last row of zeros, since it does not depend on the last weight as well as by deleting the first column, since the first value of \mathbf{z}_1^* is not a common value of the two subsamples. The gradient $d\zeta_2^*/d\mathbf{w}$ is obtained from the gradient $d\mathbf{z}_2^*/d\mathbf{w}_2$ by extending the latter by a first row of zeros, since it does not depend on the first weight as well as by deleting the last column, since the last value of \mathbf{z}_2^* is not a common value of the two subsamples. This completes the first step of the LS filter optimization.

In the second step, we calculate the constrained LS filter given the optimum weights \mathbf{w}^* obtained from the first step, i. e.,

$$\begin{aligned} & \min_{\{\mathbf{x}\}} \{F_1^*(\mathbf{x}) \equiv (\mathbf{y} - \mathbf{x})' \mathbf{W}^* (\mathbf{y} - \mathbf{x})\} \\ & \text{subject to} \\ & \quad \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \\ & \quad F_2(\mathbf{x}) \equiv \mathbf{x}' \mathbf{G} \mathbf{x} \leq \mathbf{r}' \mathbf{G} \mathbf{r} \end{aligned} \quad (3-14)$$

where \mathbf{W}^* is the diagonal matrix of the optimum weights, i. e., $\mathbf{W}^* = \text{diag}(\mathbf{w}^*)$. Since a closed-form solution of equation (3-14) does not exist, we use an optimization algorithm, given the start values of equation (1-6) and the gradients of the two functions F_1^* and F_2 . The former is given by

$$\frac{dF_1^*}{d\mathbf{x}} = 2 \mathbf{W}^* (\mathbf{y} - \mathbf{x}) \quad (3-15)$$

and the latter by equation (1-4).

In the absence of lower and upper bounds, the results for the example considered are shown in Figures 3a and 3b for four overlapping samples of equal size as well as for the tolerances $\epsilon = 2\%$. At both boundaries of the sample period, the divergence is considerably reduced when compared with the unweighted constrained LS filter of Figure 2.

4 A comparison of the LS filter with the HP filter

In order to compare the constrained LS filter with the HP filter, we introduce the $n \times n$ matrix of second differences, denoted as $\hat{\mathbf{E}}$ that is,

$$\hat{\mathbf{E}} = \begin{bmatrix} +1 & -2 & +1 & 0 & \cdots & 0 \\ 0 & +1 & -2 & +1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & +1 & -2 & +1 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \quad (4-1)$$

Let $\hat{\mathbf{G}}$ denote the $n \times n$ symmetric and positive definite matrix as follows

$$\hat{\mathbf{G}} = \hat{\mathbf{E}}' \hat{\mathbf{E}}, \quad (4-2)$$

then all the results for the LS filter carry over to the second differences upon replacing \mathbf{G} with $\hat{\mathbf{G}}$. In the absence of lower and upper bounds, the result for the example considered is shown in Figure 4a for the four overlapping samples considered in this paper.

The HP filter solves the minimization of the unconstrained LS filter given second differences, $w_0 \equiv 1$ and $(1 - w_0) \equiv \lambda$, that is,

$$\min_{\{\mathbf{x}\}} \{F_{\text{HP}}(\mathbf{x}) \equiv F_1(\mathbf{x}) + \lambda F_2(\mathbf{x})\}, \quad F_2(\mathbf{x}) \equiv \mathbf{x}' \hat{\mathbf{G}} \mathbf{x}, \quad 0 \leq \lambda < \infty, \quad (4-3)$$

where F_1 is given in equation (1-1) and λ is a penalty weight. Hodrick and Prescott [1997] suggest for monthly data a value of $\lambda_m = 14,400$. The solution of the above minimization is given by

$$\mathbf{x}_{\text{HP}}^* = [\mathbf{I} + \lambda \hat{\mathbf{G}}]^{-1} \mathbf{y} \quad (4-4)$$

We wish to find that penalty weight λ such that the HP filter mimics the second-difference unweighted constrained LS filter, denoted by $\hat{\mathbf{x}}^*$, obtained from equation (2-1) upon replacing \mathbf{G} with $\hat{\mathbf{G}}$. The objective is to minimize the deviation of the value of the penalty

weight from the suggested value, λ_m , subject to the condition that the squared differences of the HP filter and the optimum constrained LS filter is zero.⁵

$$\begin{aligned} & \min_{\{\lambda\}} \{H_1(\lambda) \equiv (\lambda - \lambda_m)^2\} \\ & \text{subject to} \\ & H_2(\lambda) \equiv (\mathbf{x}_{\text{HP}}^* - \hat{\mathbf{x}}^*)' \mathbf{I} (\mathbf{x}_{\text{HP}}^* - \hat{\mathbf{x}}^*) = 0 \end{aligned} \quad (4-5)$$

where \mathbf{x}_{HP}^* is given by equation (4-4). Since a closed-form solution of equation (4-5) does not exist, we use an optimization algorithm, given the gradients of the two functions H_1 and H_2 . The former can be stated as

$$\frac{dH_1}{d\lambda} = 2(\lambda - \lambda_m) \quad (4-6)$$

and the latter as (Goldberger, 1964)

$$\frac{dH_2}{d\lambda} = -2 \mathbf{y}' [\mathbf{I} + \lambda \hat{\mathbf{G}}]^{-1} \hat{\mathbf{G}} [\mathbf{I} + \lambda \hat{\mathbf{G}}]^{-1} (\mathbf{x}_{\text{HP}}^* - \hat{\mathbf{x}}^*) \quad (4-7)$$

Again, \mathbf{x}_{HP}^* is given by equation (4-4). Solving equations (4-5) - (4-7) for the example considered, we find for the penalty weights of the four overlapping samples considered $\lambda = \{39.8, 39.6, 42.1, 41.6\}$. The correspondig HP filters are shown in Figure 4b.

In Figure 5, we compare the optimized LS filter obtained from the third section and the unweighted constrained LS filter obtained from the second section with the HP filter which mimics the second-difference unweighted constrained LS filter. An inspection of Figure 5 shows that the HP filter exhibits a much larger variation for both in-the-sample time points and boundary points than the two LS filters mentioned above.

References

- Abramowitz, Milton and Irene A. Stegun [1972]: *Handbook of Mathematical Functions*, New York: Dover Publications, Inc., tenth edition.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross [1985]: "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53: 385 – 407.
- Goldberger, Arthur S. [1964]: *Econometric Theory*, New York: John Wiley & Sons, Inc.
- Hodrick, Robert J. and Edward C. Prescott [1997]: "Postwar U.S. Business Cycle: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29 (1): 1 – 16.

⁵ The suggested value of the penalty weight could also be set equal to zero.

The one-month inflation rate from January 1999 to October 2007

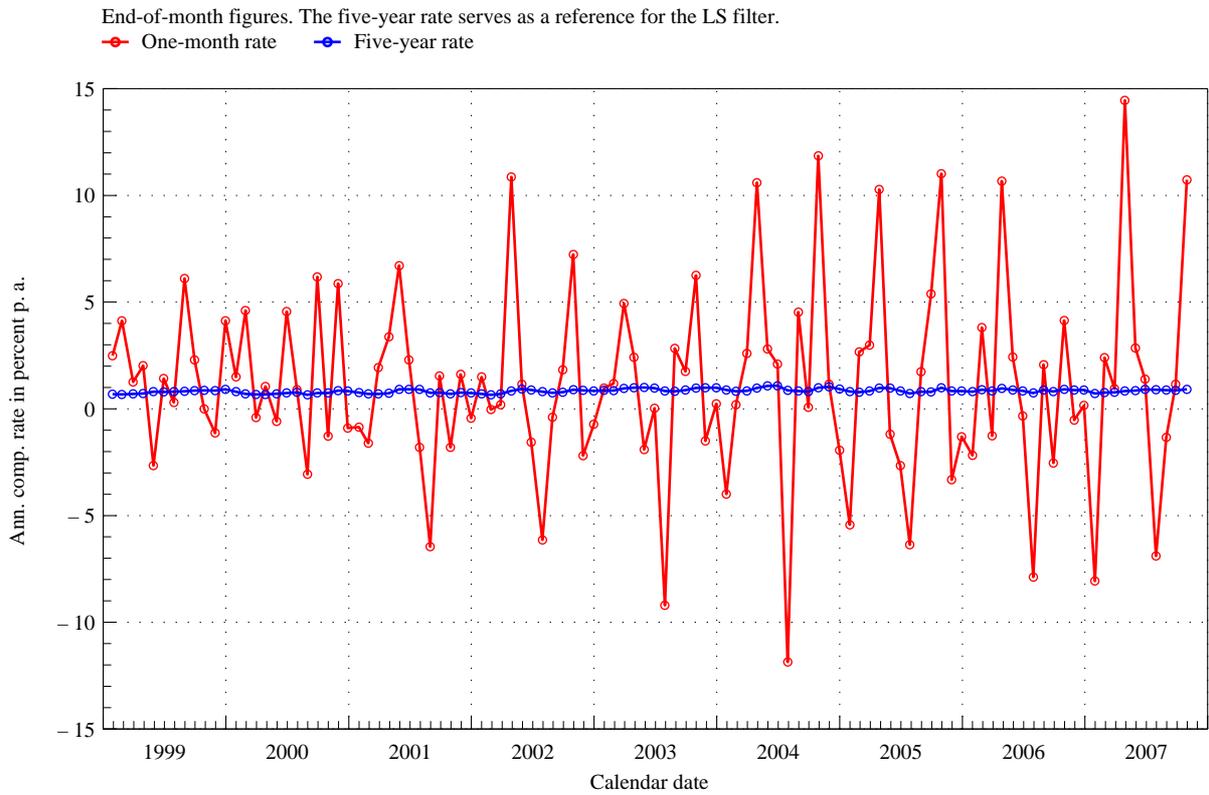


Figure 1

Unweighted constrained LS filter of the one-month inflation rate for four overlapping samples

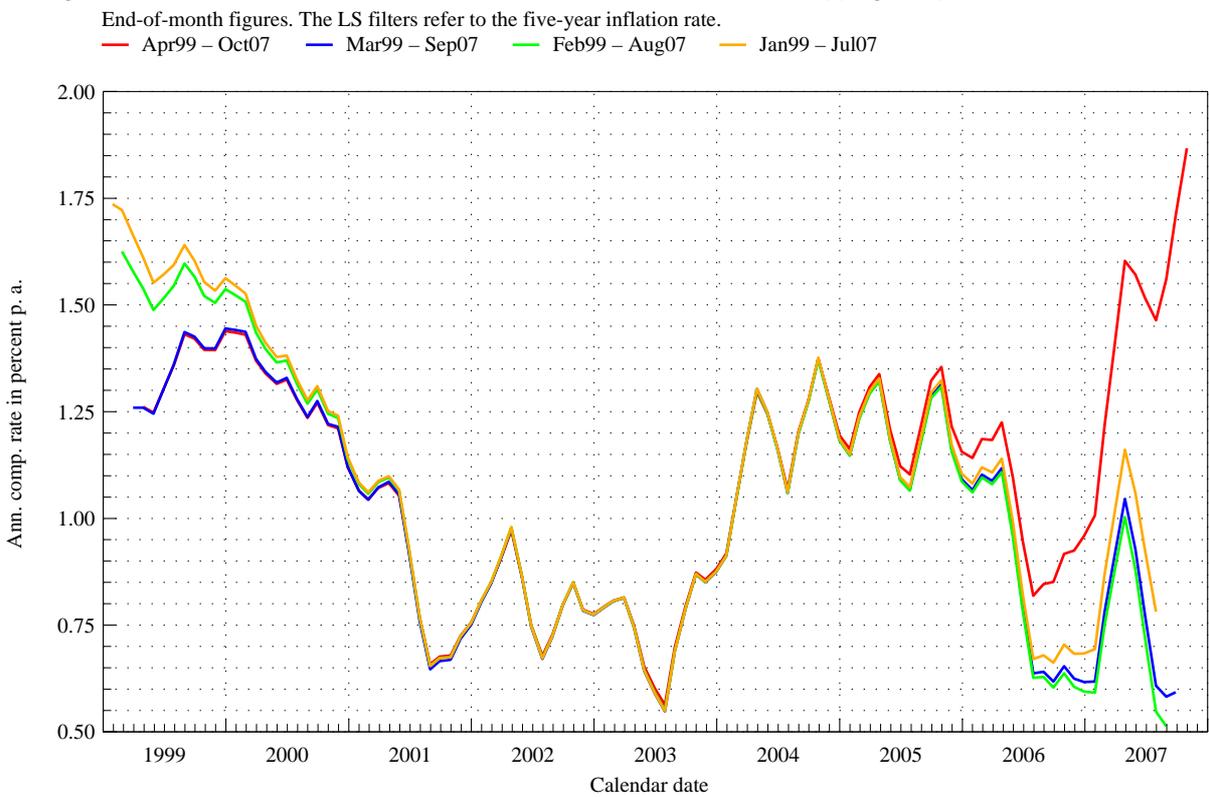


Figure 2

Optimized constrained LS filter of the one-month inflation rate for four overlapping samples

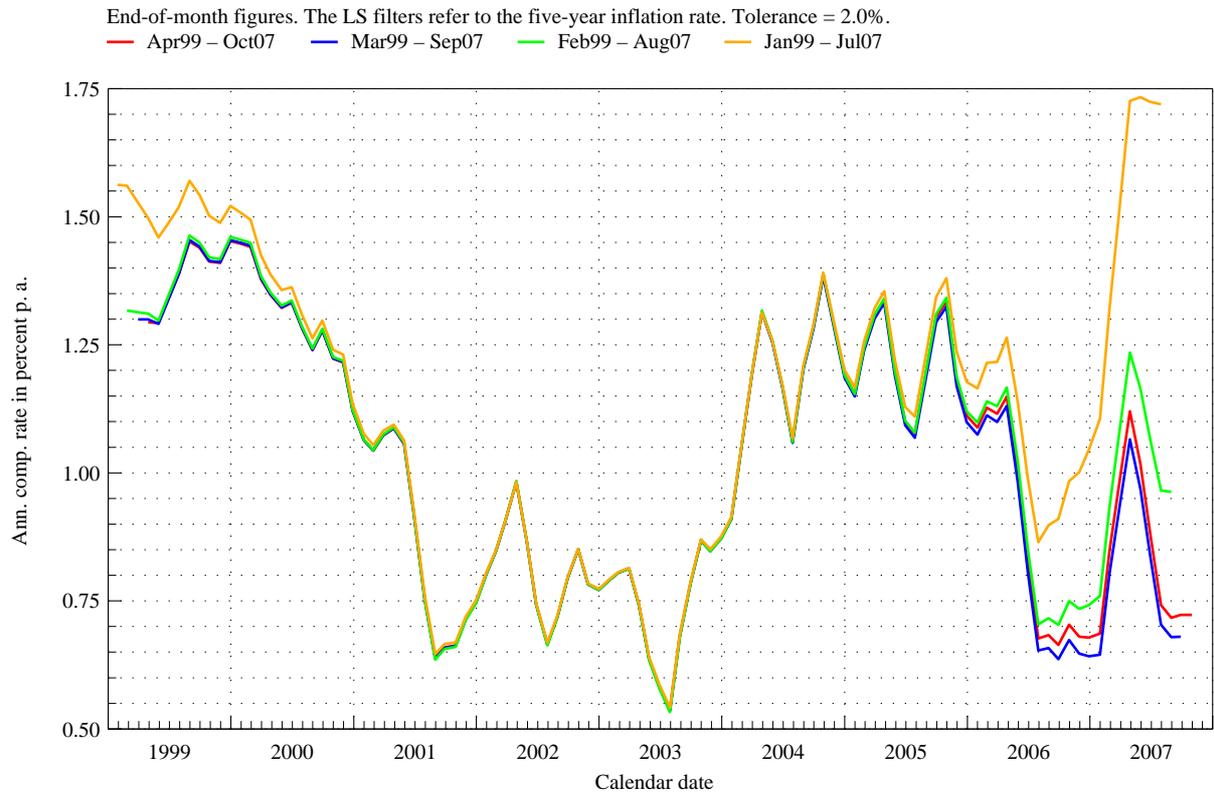


Figure 3a

Optimized weights of the constrained LS filter of the one-month inflation rate for four overlapping samples

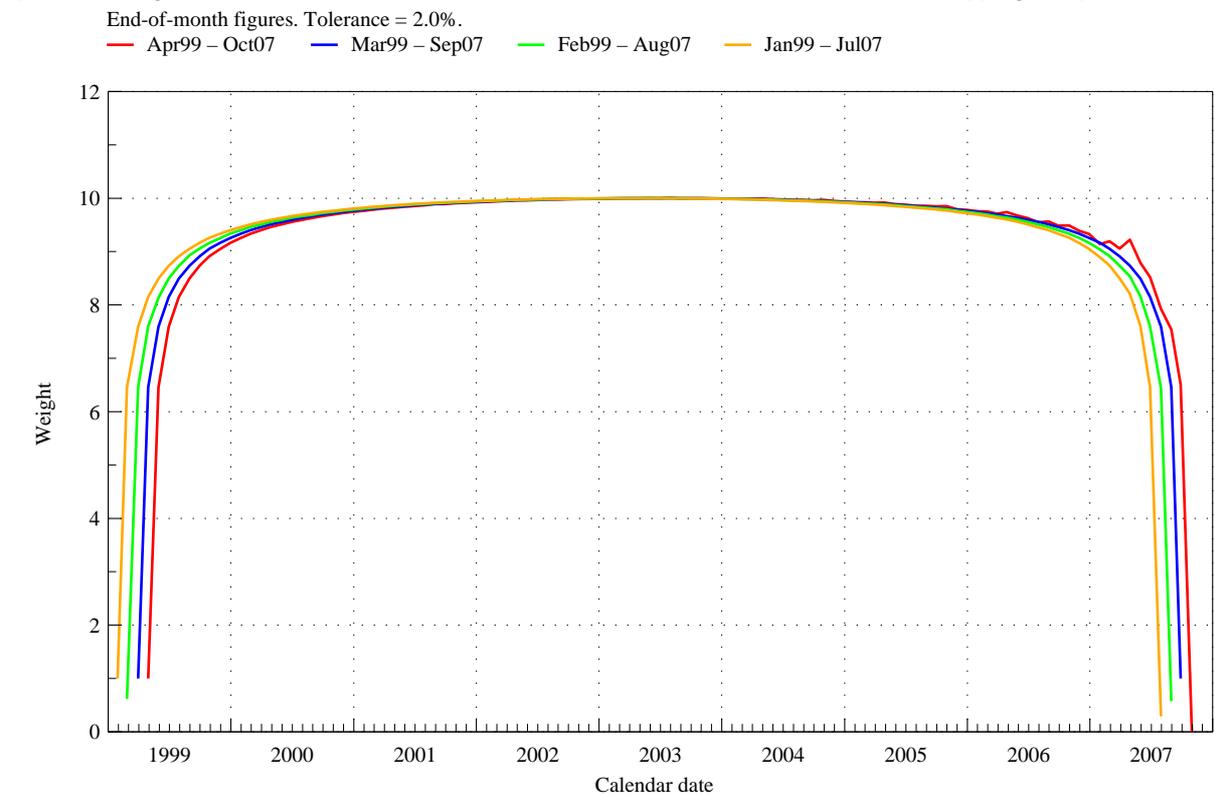


Figure 3b

Second-difference unweighted constrained LS filter of the one-month inflation rate for four overlapping samples

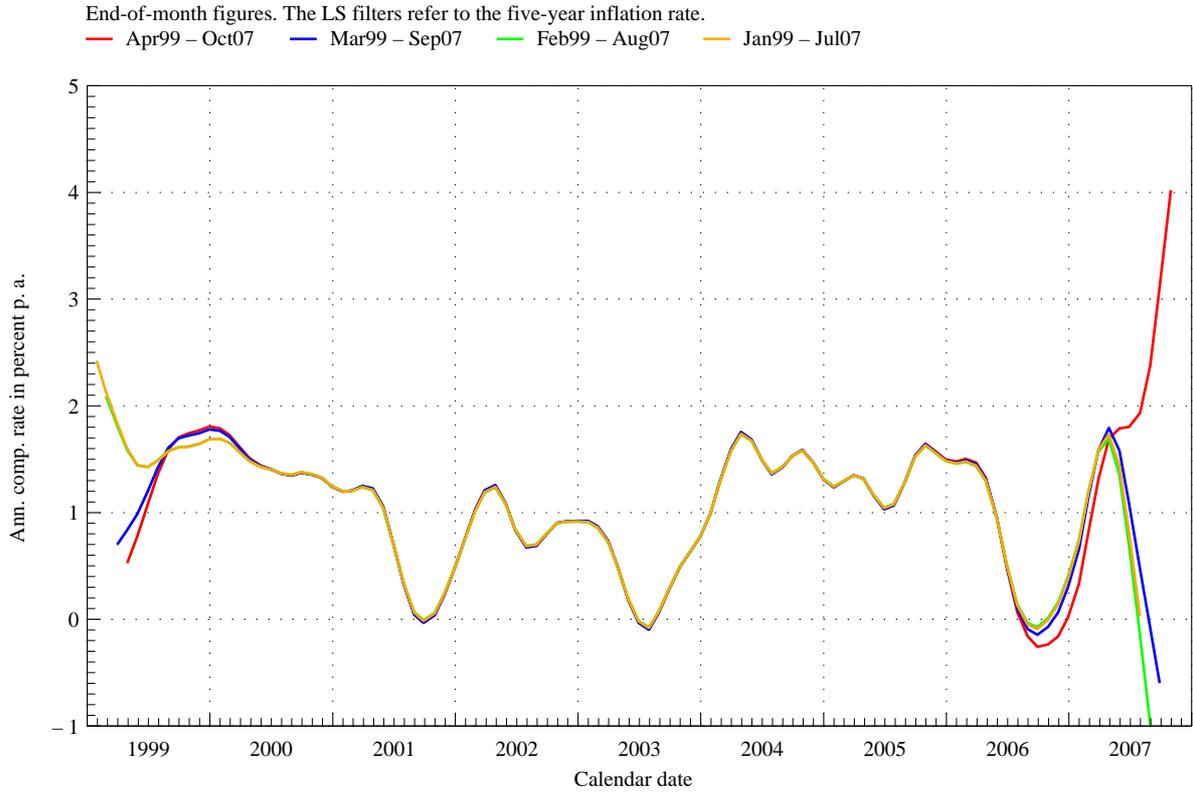


Figure 4a

HP filter of the one-month inflation rate for four overlapping samples

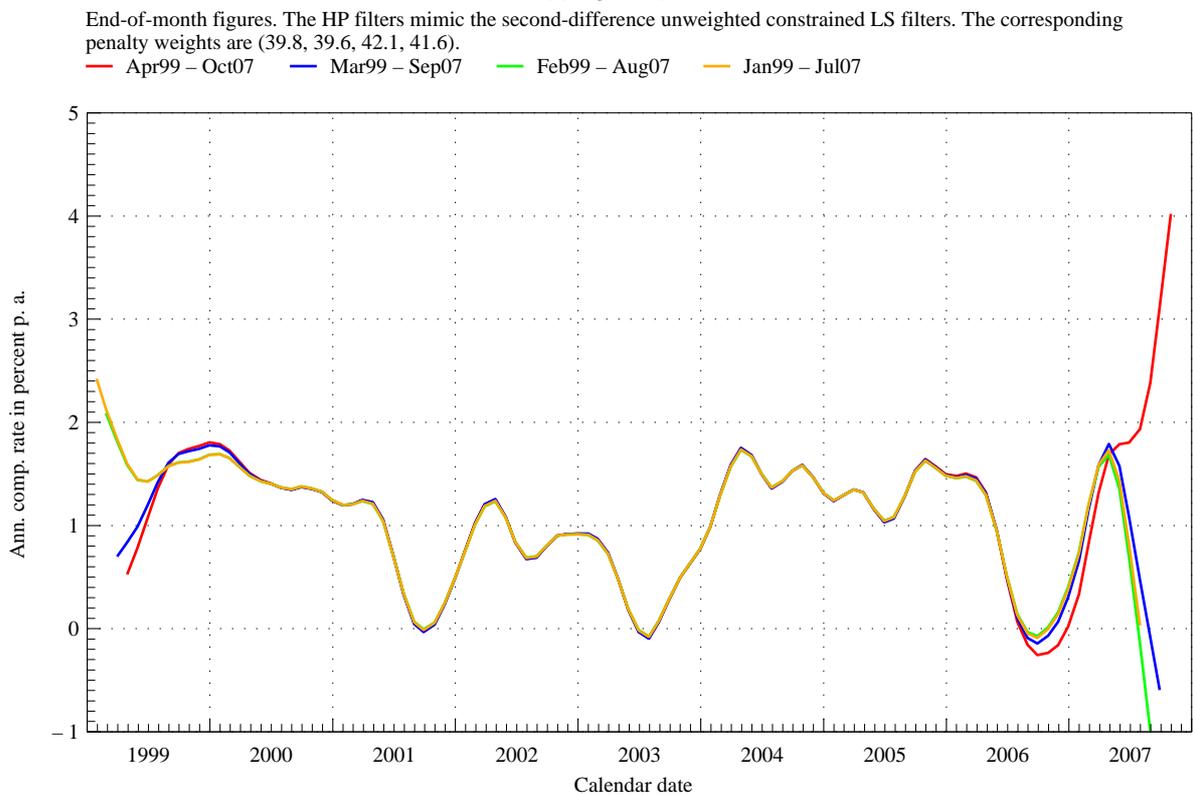


Figure 4b

Comparison of the LS filter with the HP filter

End-of-month figures. The five-year inflation rate serves as a reference for the LS filter. The optimized LS filter has a tolerance of 2.0%. The HP filter mimics the second-difference unweighted constrained LS filter with a penalty weight of 39.8. All filters refer to the period Apr99 – Oct07.

— Optimized LS filter — Unweighted LS filter — HP filter

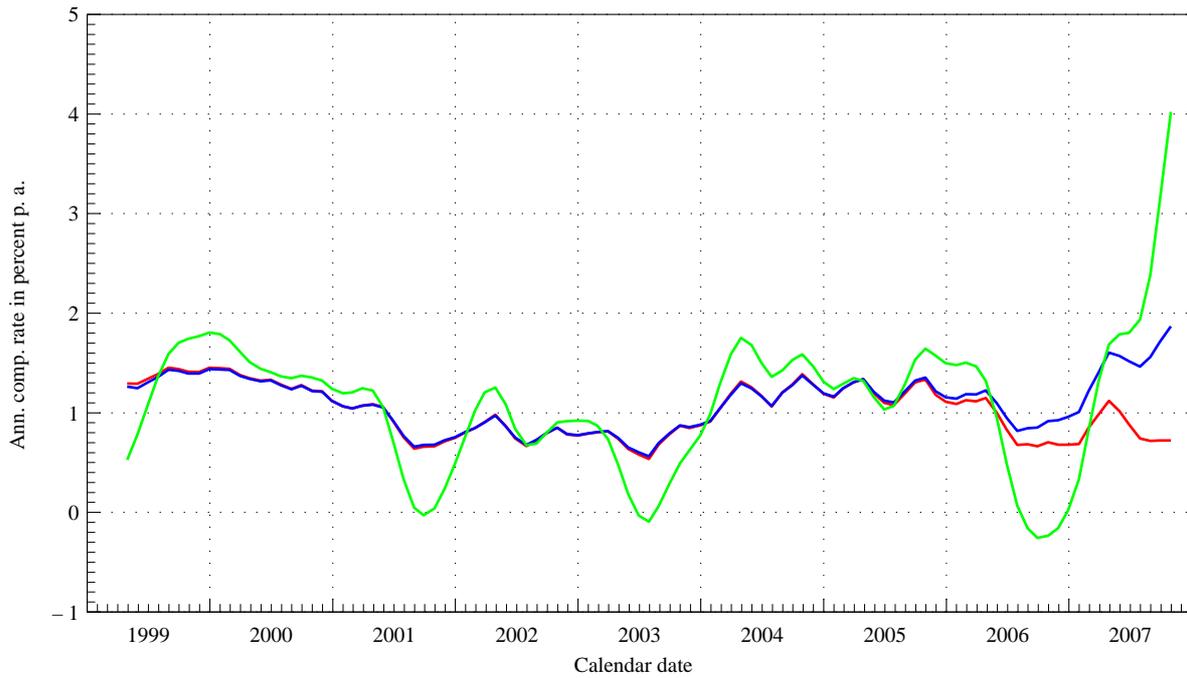


Figure 5