Journal of Modern Applied Statistical Methods

Volume 9 | Issue 1 Article 15

5-1-2010

Median-Unbiased Optimal Smoothing and Trend Extraction

Dimitrios D. Thomakos
University of Peloponnese, Tripolis, Greece, thomakos@uop.gr

Journal of Modern Applied Statistical Methods
May 2010, Vol. 9, No. 1, 144-159

1538 – 9472/10/$95.00

Median-Unbiased Optimal Smoothing and Trend Extraction

Dimitrios D. Thomakos
University of Peloponnese,
Tripolis, Greece

The problem of smoothing a time series for extracting its low frequency characteristics, collectively called its trend, is considered. A competitive approach is proposed and compared with existing methods in choosing the optimal degree of smoothing based on the distribution of the residuals from the smooth trend.

Key words: Local linear, moving average, singular spectrum analysis, smoothing, splines, time series, trend extraction.

Introduction

A fundamental problem in time series analysis is smoothing a realization and extracting its low-frequency characteristics, collectively called its trend. In the process of solving this problem a practitioner is faced with three underlying sub-problems: (a) to define the nature of the trend (e.g., deterministic or stochastic) and its perceived degree of smoothness, (b) to decide on a particular class of models to use (e.g., polynomial or non-parametric approximations), and (c) to select, usually with a data-based approach (e.g., cross-validation) the degree of approximation (or smoothness) that will enable accurate extraction of the required trend features. A large amount of literature exists which deals with these problems and includes various proposed methods for addressing them. Although it is not possible to review this literature here; many related references can be found in books and monographs, such as, Härdle (1990), Fan and Gijbels (1996), Hart (1997), Golyandina, et al. (2001) and Fan and Yao (2003).

Dimitrios D. Thomakos is a Professor in the Department of Economics at the School of Management and Economics. He is also a Senior Fellow at the Rimini Center for Economic Analysis in Rimini, Italy. Email: thomakos@uop.gr.

The methods depend on various assumptions about the data generating process (DGP) itself and its stability over time. However, in many applications one does not know or is not willing to make assumptions about the structure of the DGP and, consequently, is lead to use an approach unrelated to such specific assumptions. Examples include moving average (MA) smoothing, singular spectrum analysis (SSA) smoothing and all the known forms of non-parametric smoothing, like smoothing splines (SS) and local linear (LL) smoothers. This choice of a non-parametric approximation usually takes care of problem (b), and partially (c) if methods such as cross-validation or plug-in bandwidths are used.

As for problem (a), it is usually the case that the nature of the trend that one wants to extract is application-specific, as is its perceived degree of smoothness. However, some characteristics exist that are commonly accepted about the notion of a trend, such as: (i) it has most of its power concentrated in (a band of) the lower frequencies of the spectrum, (ii) it is more smooth (less volatile) than the actual observations, (iii) it reflects the central tendency of the process, and (iv) the observations are usually located in clusters above or below the trend component.

Problem (c) is thus left, i.e. that of selecting the appropriate optimal degree of smoothing the observations for extracting the trend component. In the context of non-parametric methods, such as SS and LL, the choice of the degree of smoothing is guided by the bias-variance trade-off and a proximity criterion - such as the mean squared error (MSE) or the integrated mean squared error - is minimized directly or by variants of cross-validation/plug-in methods. However, such criteria are invariably linked to the notion of fit (of various degrees) to the observations themselves not to the notion of an underlying trend. This runs contrary to the notion of the trend that passes through the center of the clusters of observations without tracking all their swings. In addition, for methods such as MA or SSA there are no formal procedures for selecting the degree of smoothing; the results of the proposed methodology can be applied in making such selections to these two methods as will be illustrated.

Methodology

Consider a stochastic process $\{X_t\}_{t\in Z}$ and assume that a realization of size *n* from this process is available, for example, $\{x_t\}_{t=1}^{n}$. The problem is how much to smooth the realization so as to successfully extract the low frequency characteristics, or the trend, of the process. No assumptions are made as to whether the trend is deterministic or stochastic. Such smoothing will lead to an additive decomposition of the form:

$$x_t = s_t^k + u_t^k \tag{1}$$

where s_t^k is the estimated smoothed component (the trend) of the series, that depends on a smoothing parameter *k*, and u_t^k is the estimated residual that also depends on *k*. Note that the above decomposition is not taken as the data generating process; rather it is the result of the smoothing operation. In particular, u_t^k is not assumed to be a realization from a true error process acting on X_t. As such, the representation of equation (1) has applicability both in cases where a deterministic slowly varying function of time exists and where $s_t^k = g(t/n)$ independent of *k*, and in cases where it does not, for example, in the context of a financial time series that possibly follows a random walk.

The way the residuals u_t^k are distributed is important in understanding whether a component that roughly corresponds to the characteristics (i) to (iv) attributed to the trend of a realization has been successfully extracted. First, recall that any smoothing operation that successfully extracts a measure of central tendency leads to residuals with an approximately zero mean. It does not, however, necessarily lead to residuals that have zero median, so as to have a residual distribution where equal probability is placed in observing positive (above the trend) and negative (below the trend) residuals.

This probabilistic symmetry of the residuals should be important because an extracted trend cannot possibly be accurate if it leads, on average, to more positive than negative residuals (or vice versa). In such a case the trend would be biased, either over- or under-estimating the low frequency movement of the process. If the problem of trend extraction is considered in the above context of symmetrizing the probability assigned to positive and negative residuals, it is necessary to look for a measure different than the MSE. A plausible way of proceeding is as follows.

Let $\operatorname{sgn}(x) = I(x>0) - I(x\le 0)$ denote the sign function and note that for any continuous random variable *X*, with $F_X(\cdot)$ as its distribution function,

$$\mathsf{E}\operatorname{sgn}(X) = \mathsf{E}I(X>0) - \mathsf{E}I(X\le 0) = 1 - 2F_X(0), \tag{2}$$

the absolute value of the expected sign of *X*, $|\mathsf{E}\operatorname{sgn}(X)|$, is symmetric around $F_X(0) = 0.5$ where it attains its unique minimum. It therefore follows that if the distribution of *X* is symmetric around zero (i.e., has a zero median) then $|\mathsf{E}\operatorname{sgn}(X)|$ is minimized.

This can be adapted into a smoothing context and the absolute value of the expected sign of the residuals u_t^k can serve as the objective function that should be minimized in choosing the degree of smoothing. Essentially this amounts to choosing the degree of smoothing so as to assign roughly equal probability to positive and negative residuals in accordance with characteristics (i) to (iv). This leads to consideration of the following:

$$\left|\mathsf{E}\,\mathrm{sgn}\left(u_t^k\right)\right| = \left|1 - 2F_u^k(0)\right| \tag{3}$$

where $F_u^k(\cdot)$ denotes the distribution function of the residuals. As noted, this function is minimized when $F_u^k(0) = 0.5 \Leftrightarrow u_{(0.5)}^k = 0$, that is, when the residuals are made to have a zero median. The trend component which will correspond to such residuals can now be called a median-unbiased trend.

To practically implement this idea consider the empirical version of equation (3) which can be estimated in two equivalent ways as follows:

$$\begin{aligned}\left|\mathrm{MRS}(k)\right| &\doteq \left|n^{-1}\sum\nolimits_{t=1}^{n}\mathrm{sgn}\left(u_t^k\right)\right| \\ &= \left|1-(2n)^{-1}\sum\nolimits_{t=1}^{n} I\left(u_t^k \le 0\right)\right|\end{aligned} \tag{4}$$

where MRS denotes the mean residual sign based on the sample of observations. As can be observed from equation (4), the MRS can be obtained either using the average sign or using the empirical distribution function evaluated at zero. The most practical way of optimizing the $\left|\mathrm{MRS}(k)\right|$ is by direct search over a grid of plausible values for the smoothing parameter. If $K \doteq \{k_{\min}, k_{\max}\}$ denotes such a grid then the optimal value k^* is given by:

$$k^* = \arg\min\nolimits_{k\in K}\left|\mathrm{MRS}(k)\right| \tag{5}$$

The range of grid values to consider is both problem-specific and method-specific and no general guidelines can be given. For example, if a moving average is to be used for smoothing, then k takes only integer values; if a kernel smoother is to be used then k takes real values – possibly in a pilot interval. To overcome this potential shortcoming one can alternatively consider using data-dependent, sub-sampling approaches. One variant of such a sub-sampling approach could be as follows:

1. Split the observations into M non-overlapping sections each of equal length $m \le n/2$, for $j = 1,\dots,M$, with $m \to \infty$ as $n \to \infty$ and $m/n \to c$ for some constant c.
2. Select a range of plausible values for each section, for example K_j.
3. Compute the optimal value of the smoothing parameter for each section, for example k_j^*.
4. Select the full sample optimal value of the smoothing parameter as the average of the parameters from each section, i.e., $k^* = M^{-1}\sum\nolimits_{j=1}^{M} k_j^*$.

The above is just one sub-sampling method. Alternatively, the series can be split using a sliding window of length m, thus having M overlapping sections each of length m. This alternative is not further pursued herein but is easily implementable.

Results

The above methodology was applied to simulated time series and a real time series using different smoothers: symmetric MA, SSA, LL, SS and the Kalman fixed point (KF) smoother. All methods are appropriate under different conditions for the data generating process. For the LL smoothing and SS methods the degree of smoothing selected by the present methodology was compared with the degree of smoothing selected using generalized cross-validation (GCV) and plug-in (plug) methods respectively. The SSA smoother was used as in Thomakos (2008) with an asymptotically optimal decomposition of the covariance of the process, when the process has stochastic trends. All computations reported below were performed in R.

Simulated Series

Two types of data generating process (DGP) were considered. The first is given as the sum of a deterministic, slowly varying function $g(t/n)$ and stationary errors and the second is given as the sum of a stochastic trend (a random walk) and stationary errors. Specifically, for the first DGP:

$$\text{DGP I: } x_t = g(t/n) + u_t \qquad (6)$$

with $g(t/n) = \alpha + \beta t + \sum_{j=1}^{2} \gamma_j \cos\left(\frac{2\pi\omega_j t}{n}\right)$ and with $u_t = \phi u_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. For the trend function $g(t/n)$ the critical parameters determining the degree of smoothness (and the complexity of the curve) are the frequencies $f_j = \omega_j / n$; higher values decrease smoothness - see Figure 1 for an illustration (the black line corresponds to the less smooth trend, the red (or gray) line to the more smooth trend).

For the final series x_t the critical parameters are $\left[\phi, \sigma_\varepsilon^2 / (1-\phi^2)\right]$, the persistence and the variance of the error term; higher values make it more difficult to separate the trend from the errors. In the end, consider the following combinations for the parameters:

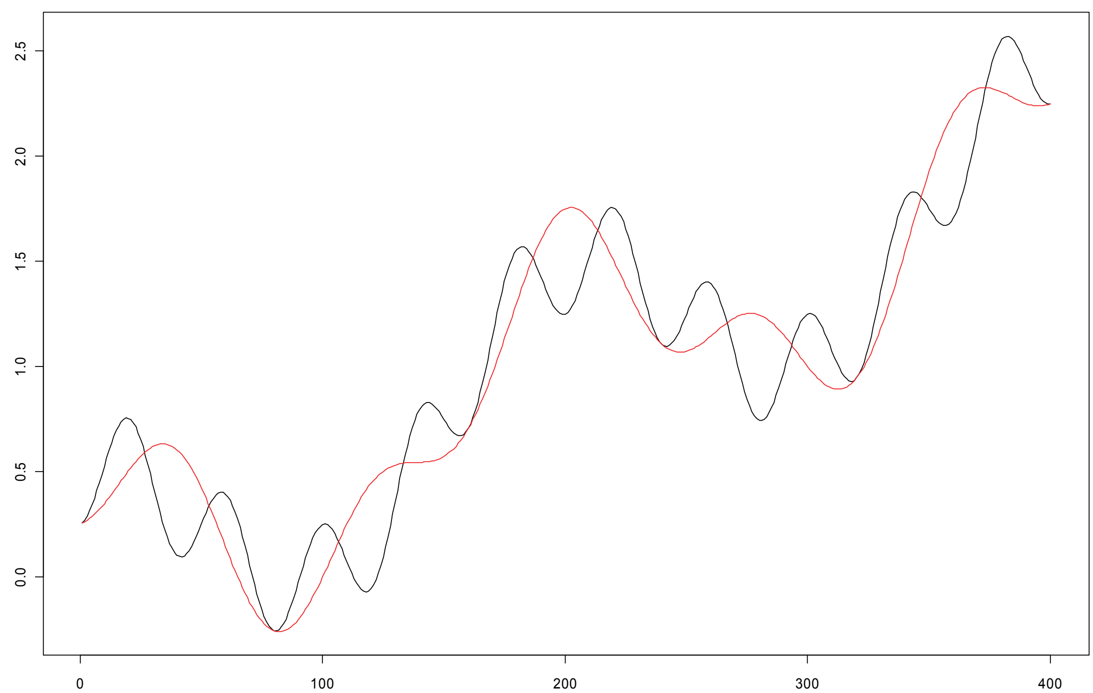
$$\begin{aligned} &\alpha = 0, \beta = 2, \gamma_1 = 0.50, \gamma_2 = -0.25, \\ &\omega_1 = 2, \omega_2 = \{5,10\}, \phi = \{0.0, 0.8\}, . \\ &\sigma_\varepsilon^2 = 0.2^2 \end{aligned}$$

For the second DGP consider the well known form of signal-plus-noise or local level model as:

$$\text{DGP II: } x_t = g(\alpha, S_t) + u_t \qquad (8)$$

with $g(\alpha, S_t) = \alpha + S_t$, where α is the drift parameter, $S_t = \sum_{j=1}^{t} \varepsilon_j$ is the random walk component of the series with normally

Figure 1: The Smooth Trend Functions from DGP I of Equation (6)



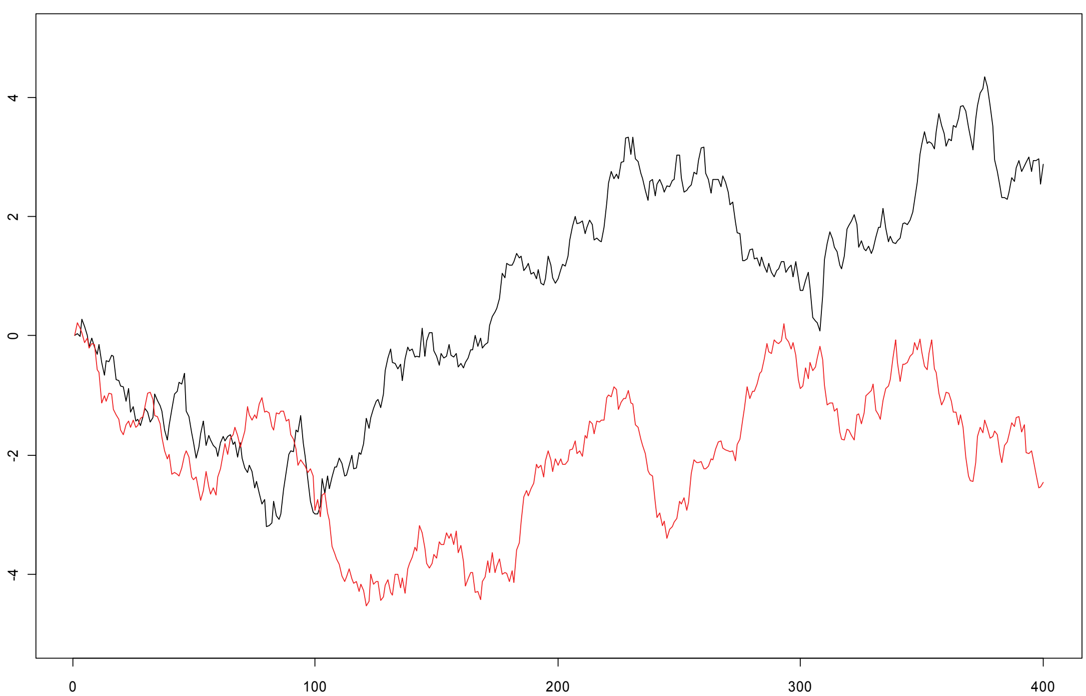
distributed errors $\varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)$, and where u_t are the added errors that have either a normal or a t-distribution, that is, $u_t \sim N\left(0,\sigma_u^2\right)$ or t $u_t \sim t_{(6)}$. The drift parameter is set to $\alpha = 0.1$, the variance term of ε_t is set to $\sigma_\varepsilon^2 = 0.2^2$ and the variance of the normally distributed u_t is set to $\sigma_u^2 = 0.6^2$ (the later corresponds to a 1:3 signal-to-noise ratio). Typical sequences from the DGP of equation (8) are shown in Figure 2 (the black (upper) line corresponds to normally distributed additive errors, the red (lower) line to additive *t*-distributed errors).

From each DGP, r = 1, 2 … 400 realizations of sizes n = {200, 400} were simulated and for each realization the full sample and the sub-sampling approach was used, the latter with m = {50, 100} for the corresponding sample sizes, to compute the optimal value of the smoothing parameters of each method. The ranges of plausible values for minimizing the MRS were set to the following:

- For the symmetric MA and SSA methods that use integer values: $K = \{2k+1 \mid k = 1,2,\ldots,11,12\}$
- For the local linear smoothing that uses real values for the bandwidth:
$$K = \left\{1.5^{k-12} s_x \mid k = 1,2,\ldots,11,12\right\},$$
where s_x denotes the standard deviation of the data.
- For the smoothing splines that use real values for the smoothing parameter $K = \{k \mid k = 0.00, 0.14, \ldots 1, .36, 1.50\}$, a sequence of 12 values in the interval [0.0, 1.50].

With the selected k^*, as computed either with the full sample or the sub-sampling approach, the mean absolute deviation of the true trend component from the estimated trend component is computed for each replication, that is $m_r^{k^*} = n^{-1}\sum_{t=1}^{n}\left|g_r(\cdot) - s_{t,r}^{k^*}\right|$. Finally, the

Figure 2: Sample Realization of Stochastic Trend from DGP II of Equation (8)



average, $m^{k^*} = 400^{-1}\sum_{r=1}^{400} m_r^{k^*}$, was computed as well as the optimal values of k^* from all 400 replications (note that the reported replication averages for integer k^* will not necessarily be odd numbers). These measures are reported in Tables 1 and 2 which show the results on the simulations for DGP I, and in Table 3 which shows the results for DGP II.

Discussion of Simulation Results: DGP I

For the smaller sample size of n = 200 (see Table 1), the discussion can be separated into two cases: one for $\phi = 0$ and the other for $\phi = 0.8$; for the first case also note some small differences depending on the value of ω_2. Thus, for the parameter combination $\phi = 0, \omega_2 = 5$ the performance of the sub-sampling approach is improving the average accuracy in extracting the trend for the MA and SS methods. There is no change for the local linear smoother. Note that only the moving average coupled with sub-sampling performs on par with the GCV-based and plug-in approaches but this is an important result: the smoothing spline and local linear methods have their own approaches (GCV and plug-in) for selecting the degree of smoothing while the for a moving average there is no such existing method.

For the parameter combination $\phi = 0, \omega_2 = 10$ however the results are much less satisfactory since no alternative beats the GCV-based and plug-in-based approaches. Turning next to the parameter combinations where $\phi = 0.8$ a much improved picture results in terms of the performance of the proposed methodology and the use of moving averages.

Table 1: Average Absolute Deviation of True from Estimated Trend & Optimal Degrees of Smoothing; Simulations from DGP I and Sample Size n = 200

ϕ = 0 and ω_2 = 5

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.09	0.06	0.13	0.08	0.05	0.13	0.16	0.05
k^*	10	8	0.48	0.33	0.58	0.09	0.09	0.02

ϕ = 0.8 and ω_2 = 5

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.19	0.19	0.24	0.23	0.24	0.21	0.21	0.20
k^*	8	7	0.45	0.31	0.22	0.08	0.08	0.01

ϕ = 0 and ω_2 = 10

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.13	0.13	0.14	0.09	0.06	0.17	0.17	0.07
k^*	10	8	0.50	0.30	0.44	0.11	0.07	0.01

ϕ = 0.8 and ω_2 = 10

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.21	0.21	0.24	0.23	0.24	0.22	0.22	0.21
k^*	9	7	0.44	0.32	0.22	0.08	0.07	0.01

Here the use of either the full or sub-sampling approaches coupled with a moving average produces is better (when $\omega_2 = 5$) or on par (when $\omega_2 = 10$) with the alternative methods.

When the sample size increases to n = 400 (see Table 2) further improvements are observed in performance from the use of the proposed methodology – especially from the use of moving averages. Specifically, in all four parameter combinations considered, a moving average coupled with sub-sampling performs on par or better than GCV-based and plug-in-based approaches. Note that this improvement is more pronounced in some cases and is worth elaborating about. For example, in the case where $\phi = 0.8, \omega_2 = 5$ the moving average performs on par with the local linear smoother with plug-in selection of bandwidth; the smoothing splines do not perform as well. In the case where $\phi = 0.8$, $\omega_2 = 10$, the moving average with sub-sampling outperforms the local linear smoother. Finally, the smoothing splines with sub-sampling now perform on par with the GCV-based smoothing splines.

The results from the DGP I simulations show that the proposed methodology can be competitive to existing methods, by either: (1) assisting less sophisticated methods, such as moving averages, to perform well in smoothing and trend extraction, and/or (2) producing results using other methods, such as smoothing splines that are equivalent to the more sophisticated GCV or plug-in approaches.

Simulation Results Discussion: DGP II

Recall that the simulations of the second DGP of equation (8) do not have an underlying deterministic smooth function that serves as the trend component, but rather have a stochastic trend that is masked by additive errors. This type of DGP has a corresponding optimal smoother,

Table 2: Average Absolute Deviation of True from Estimated Trend & Optimal Degrees of Smoothing; Simulations from DGP I and Sample Size n = 400

$\phi = 0$ and $\omega_2 = 5$

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.05	0.04	0.13	0.06	0.03	0.12	0.16	0.04
k^*	11	10	0.55	0.40	0.59	0.10	0.09	0.02

$\phi = 0.8$ and $\omega_2 = 5$

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.17	0.17	0.22	0.20	0.24	0.20	0.18	0.17
k^*	10	9	0.46	0.36	0.15	0.10	0.07	0.01

$\phi = 0$ and $\omega_2 = 10$

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.10	0.07	0.14	0.07	0.05	0.15	0.17	0.05
k^*	11	10	0.58	0.38	0.45	0.11	0.07	0.01

$\phi = 0.8$ and $\omega_2 = 10$

Smoother	MA-full	MA-sub	SS-full	SS-sub	SS-GCV	LL-full	LL-sub	LL-plug
m^{k^*}	0.19	0.17	0.22	0.21	0.24	0.20	0.20	0.18
k^*	10	9	0.50	0.34	0.15	0.09	0.07	0.01

based on the state space representation of equation (8), the Kalman fixed point smoother. Results can thus be compared to this natural benchmark. Here the results are much more uniform across sample sizes and distributions and highly encouraging. For all cases considered in Table 3 there is at least one instance of either the MA or the SSA smoother, with sub-sampling, that

From the above discussion it is clear that a carefully, data-based, selected MA or SSA smoother can potentially perform as well or better than more sophisticated methods when extracting a stochastic trend from underlying additive errors. Note that the simplicity of these methods is important in the context of this discussion: they require no assumptions about the DGP of the problem to be made and can thus be applied universally.

Real Series: The U.S. GDP

An interesting series, for which the current methodology is relevant, is that of the United States real Gross Domestic Product (GDP - series GDPC96 from the Federal Reserve Bank of St. Louis online database). This analysis includes the last 200 available quarters for the years 1958 to 2008.

This series is the main economic indicator for the United States and from it the so-called output gap and the growth rate of the economy is computed. The logarithm of this series is plotted in Figure 3 which shows its salient characteristics, namely that it appears to be quite smooth and that it contains a trend component, which corresponds to the long-run (low frequency) movement of the economy.

Considerable literature exists in economics related to which type of stochastic process is best suited for describing the series.

Table 3: Average Absolute Deviation of True from Estimated Trend & Optimal Degrees of Smoothing; Simulations from DGP II and Sample Sizes $n = \{200, 400\}$

Normally Distributed Errors, n = 200

Smoother	MA-full	MA-sub	SSA-full	SSA-sub	SS-GCV	KF-full	LL-sub	LL-plug
m^{k^*}	0.24	0.22	0.22	0.21	0.21	0.20	0.32	0.21
k^*	9	7	9	7	0.56	n.a.	0.12	0.02

Normally Distributed Errors, n = 400

Smoother	MA-full	MA-sub	SSA-full	SSA-sub	SS-GCV	KF-full	LL-sub	LL-plug
m^{k^*}	0.25	0.24	0.22	0.21	0.20	0.20	0.52	0.22
k^*	11	9	11	9	0.42	n.a.	0.19	0.02

t-Distributed Errors, n = 200

Smoother	MA-full	MA-sub	SSA-full	SSA-sub	SS-GCV	KF-full	LL-sub	LL-plug
m^{k^*}	0.33	0.32	0.36	0.34	0.30	0.30	0.37	0.30
k^*	9	7	8	7	0.75	n.a.	0.13	0.04

t-Distributed Errors, n = 400

Smoother	MA-full	MA-sub	SSA-full	SSA-sub	SS-GCV	KF-full	LL-sub	LL-plug
m^{k^*}	0.33	0.31	0.35	0.32	0.30	0.29	0.52	0.30
k^*	11	10	9	9	0.59	n.a.	0.18	0.03

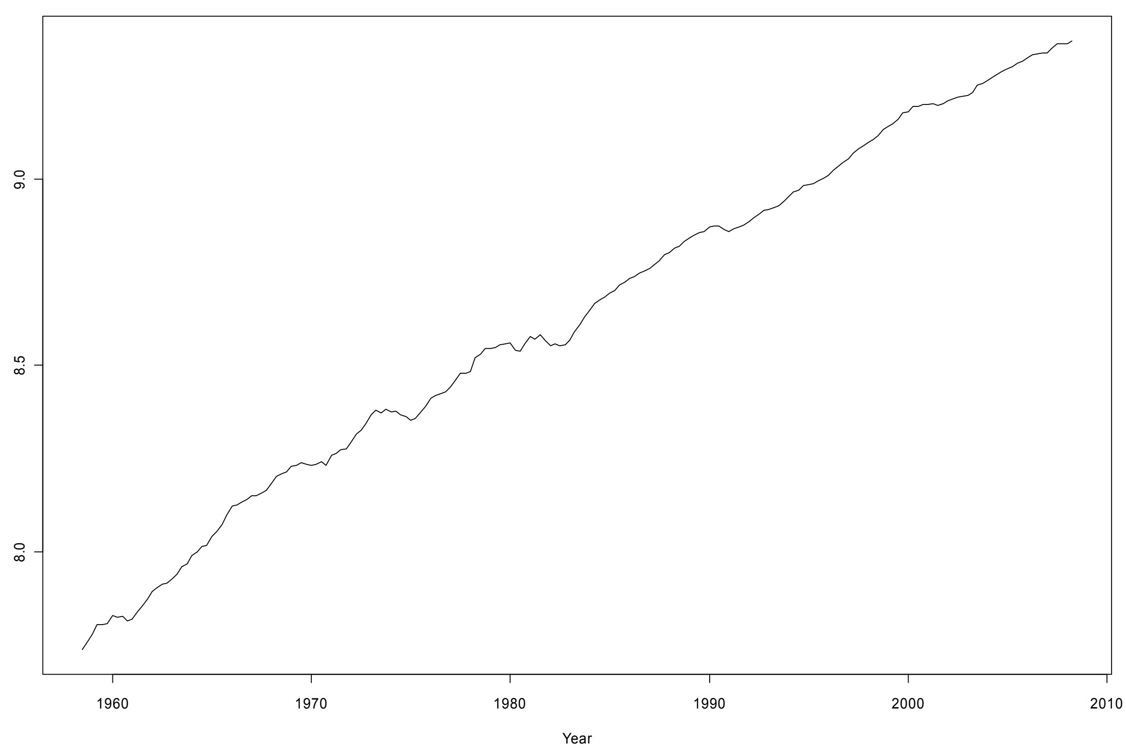
However, no claims as to which process is indeed appropriate are put forth herein. Despite the visual proximity, it is not clear if a global deterministic trend is observed or a particular manifestation of a stochastic trend $g_t\left(\alpha_t, S_t\right)$ with structural changes. No definite answer has emerged from the related literature but the consensus agrees that a deterministic linear trend will be a poor approximation both because its shape does not agree with the underlying economic intuition and because it is not expected that such a global structure will remain stable over long periods of time. Therefore alternative ways of extracting the trend component by filtering or smoothing must be considered.

The most popular smoother, in this and related macroeconomic contexts is the Hodrick and Prescott (1997) or HP smoother. Note that this smoother is only optimal under specific conditions for the DGP (see for example Dermoune, et al., 2007). Nevertheless, it is so frequently used that its merits as an accurate representation of the DGP are not further discussed. The performance of the HP smoother and its degree of smoothing being selected by various methods are compared with the performance of the other smoothers we considered previously.

The potential differences from the application of different smoothing methods in the GDP series can only be assessed indirectly because there is no true trend component with which to compare results. Thus, the residuals after smoothing - the output gap - are considered as the variable of interest on which performance comparisons can be made.

The full and sub-sampling approaches have been applied to the MA, SSA and HP smoothers. In addition, the GCV-based smoothing splines were considered along with the plug-in based local linear smoother and the HP filter with an optimally selected value for the degree of smoothness (Dermoune, et al., 2007). Denote by $u_t^{k^*,j}$ the residuals obtained

Figure 3: Natural Logarithm of the U. S. Real Gross Domestic Product, 1958 to 2008



from the j^{th} method and by $u_t^{k^*,HP-opt}$ the residuals obtained using the HP smoother with an optimally selected degree of smoothing. For each of these series we report their sample standard deviation and a Kolmogorov-Smirnov type test for the differences in the empirical cumulative distribution between $u_t^{k^*,j}$ and $u_t^{k^*,HP-opt}$. To compute the latter test the following steps are used:

1. Compute the empirical distributions of $u_t^{k^*,j}$ and $u_t^{k^*,HP-opt}$, for example, $F_n^j(u)$ and $F_n^{HP-opt}(u)$, over a grid of values, for example, $u \in U \subset R$.
2. Compute the Kolmogorov-Smirnov statistic $D_n = \sup_{u\in U} \sqrt{n}\left|F_n^j(u) - F_n^{HP-opt}(u)\right|$ for testing the equality of the underlying distributions.
3. Obtain an appropriate critical value for the test in the above step using the bootstrap – the stationary bootstrap (see Politis & Romano, 1994) was used in this study.

A number of interesting results are summarized and can be read from Table 4. Immediately it is observed that the hypothesis of equal distributions for the output gap between the HP smoother and all the other smoothers is not rejected. Therefore, in terms of the distribution of the residuals, all smoothers are essentially equivalent.

In addition there are a number of other interesting results that can be deduced from Table 4. First, note that the standard deviation of the residuals for the HP-based methods is practically the same irrespective of whether one uses the optimally selected degree of smoothing, as in Dermoune, et al. (2007), or uses the full or the sub-sampling methodology proposed herein.

Second, the MA and SSA smoothers produce residuals with larger standard deviation than the previous HP smoothers but which are on par with the standard deviation of the residuals obtained when the HP smoother is applied with the default degree of smoothing (equal to 1,600) as originally recommended by Hodrick and Prescott (1997). That value of the standard deviation was found to be 0.015. Finally, as shown in Figures 4, 5 and 6, the smoothers can be clustered together based on the standard deviations of their residuals to visualize their similarities and differences.

In Figure 4 the residuals from the three HP smoothers are plotted as in Table 4 plus the GCV-based smoothing spline smoother; it may be observed that the series are practically identical and this lends considerable support to the methodology proposed in this article as the residual series of the optimal HP smoother is able to be reproduced using both the full and sub-sampling approaches in minimizing the mean residual sign.

In Figure 5 the residuals from the MA (full and sub-sampling), the SSA (full only) and the default HP smoothers are plotted. Again a remarkable degree of closeness in the shape and magnitude of the four series is observed, especially of the moving average with sub-sampling and the default HP smoother.

Finally, Figure 6 plots the residuals from the singular spectrum analysis smoother with sub-sampling and the local linear smoother with plug-in bandwidth and, again, the series look practically identical.

Table 4: Standard Deviation of Residuals After Smoothing And Bootstrap-Based P-Value of the Kolmogorov-Smirnov Test for Equality of Distributions Between Residual Series and the Residuals from the HP Smoother with Optimally Selected Degree of Smoothness

Smoother	*HP-opt*	*HP-full*	*HP-sub*	*MA-full*	*MA-sub*	*SSA-full*	*SSA-sub*	*SS-GCV*	*LL-plug*
SD of Residuals	0.002	0.003	0.003	0.018	0.015	0.018	0.007	0.003	0.006
p-value	n.a.	0.922	0.822	0.962	0.902	0.972	0.717	0.800	0.825

Figure 4: Residuals from Three HP Smoothers and the Smoothing Spline

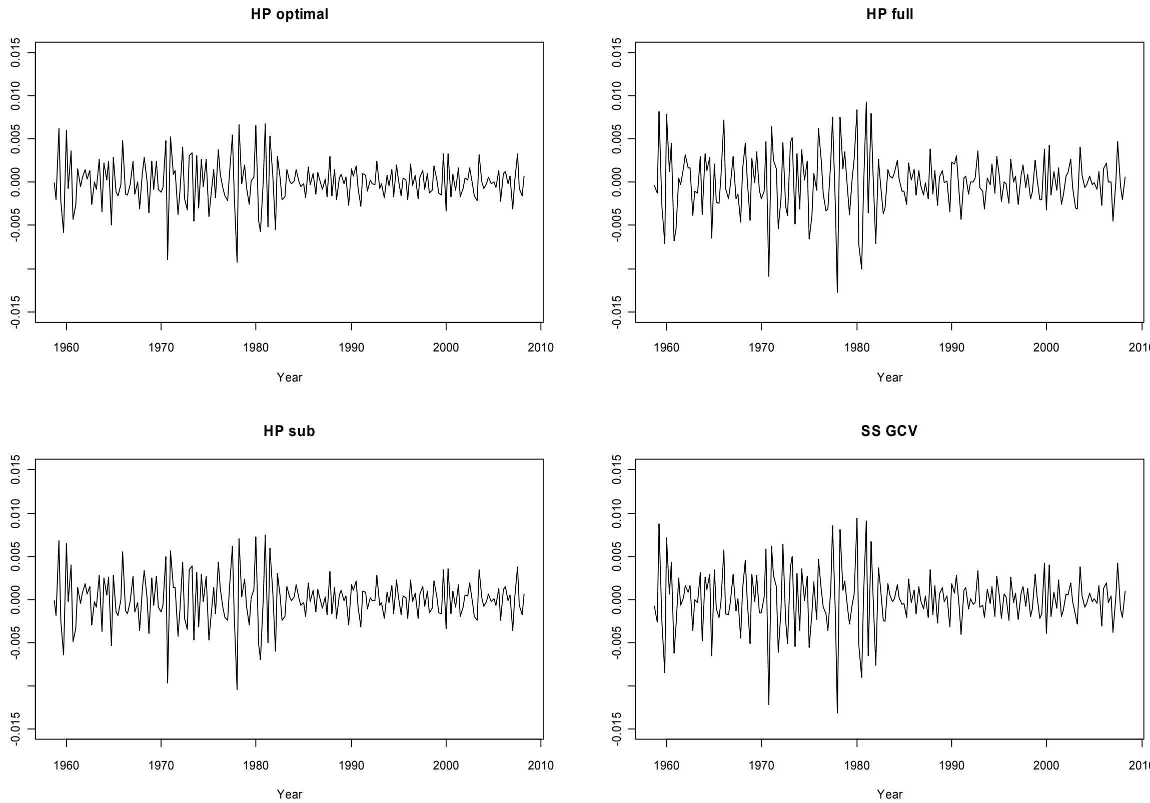


Figure 5: Residuals from MA, SSA and Default HP Smoothers

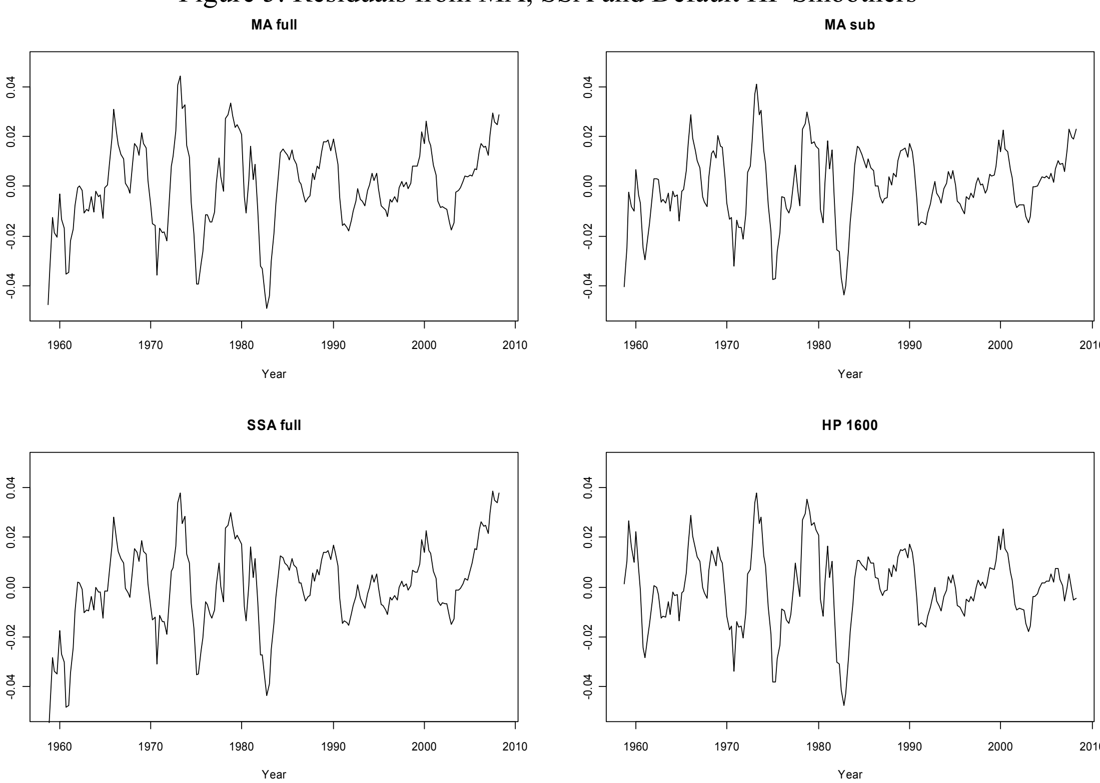
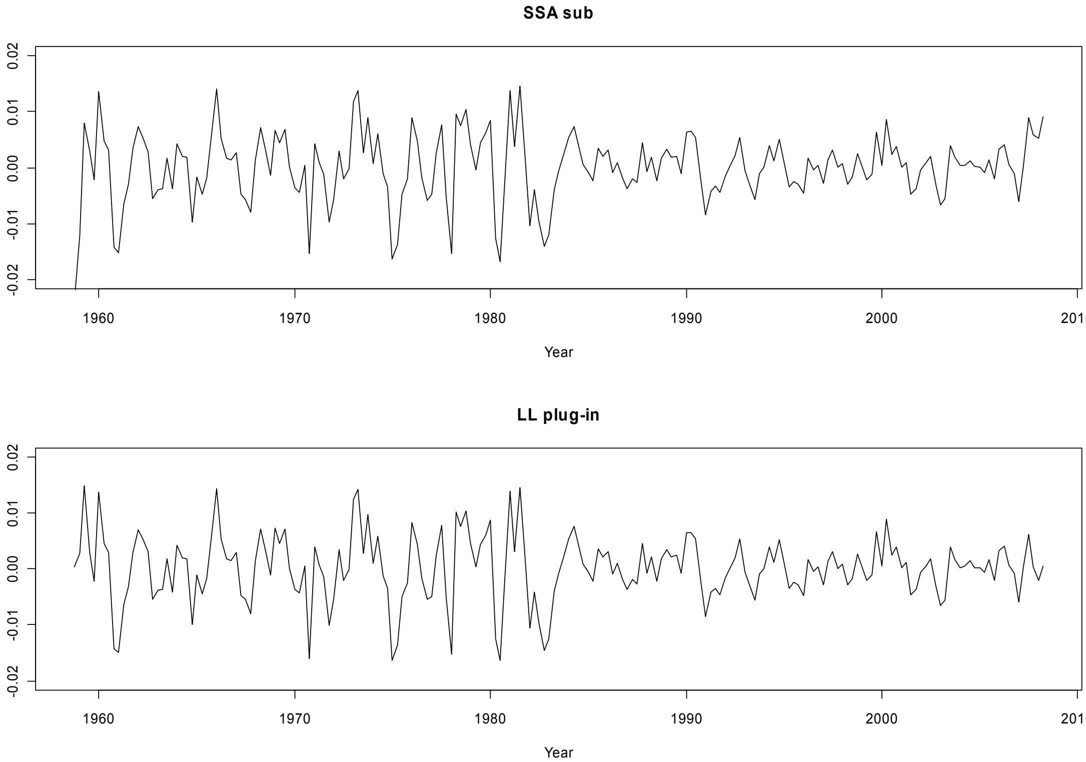


Figure 6: Residuals from Singular Spectrum Analysis and Local Linear Smoothers



It is evident from the figures that both similarities and differences exist among the smoothers and this is due to both their underlying filters and to the way the optimal degree of smoothing is selected. To explain the results consider the fact that the HP smoother is the optimal smoother for a stochastic process that is stationary in second differences. Therefore its application will necessarily lead to excess differencing if the true DGP becomes stationary after first differencing. Because the first differences of the GDP series are probably stationary (see Figure 7), then the HP smoother will remove a broader band of frequency components than the one corresponding to the trend of the series. The same holds true for the GCV-based smoothing splines smoother. To visualize this observe the shapes of the series in Figures 4 to 6; it can also be judged from the shapes of their corresponding autocorrelation or spectral density functions.

In Figures 8, 9 and 10 the spectral densities of the series that correspond to Figures 4, 5 and 6 are presented, these figures reinforce The spectral shapes in Figure 8 show that the application of the HP smoother, with optimally selected degree of smoothing, removed the power corresponding to the business cycle frequencies, corresponding from 6 to 32 quarters (see for example Christiano & Fitzgerald, 2003). Its application is thus removing not just the trend but also the business cycle component of the series. Conversely, the spectral shapes in Figures 9 and 10 are more in line with one another and with the idea of optimal smoothing for trend extraction. In all plots in these two figures the spectral densities have a single clear peak at frequencies corresponding to about 20 quarters (Figure 9) and 12 quarters (Figure 10) respectively. Both of these numbers fall within the range of the business cycles frequencies noted above. In fact, the peak of 12 quarters obtained by the smoothers in Figure 10 is almost the mid-range of the business cycles frequencies. Either the SSA smoother with sub-sampling or

Figure 7: Autocorrelation Function of Quarterly U.S. Real GDP Growth

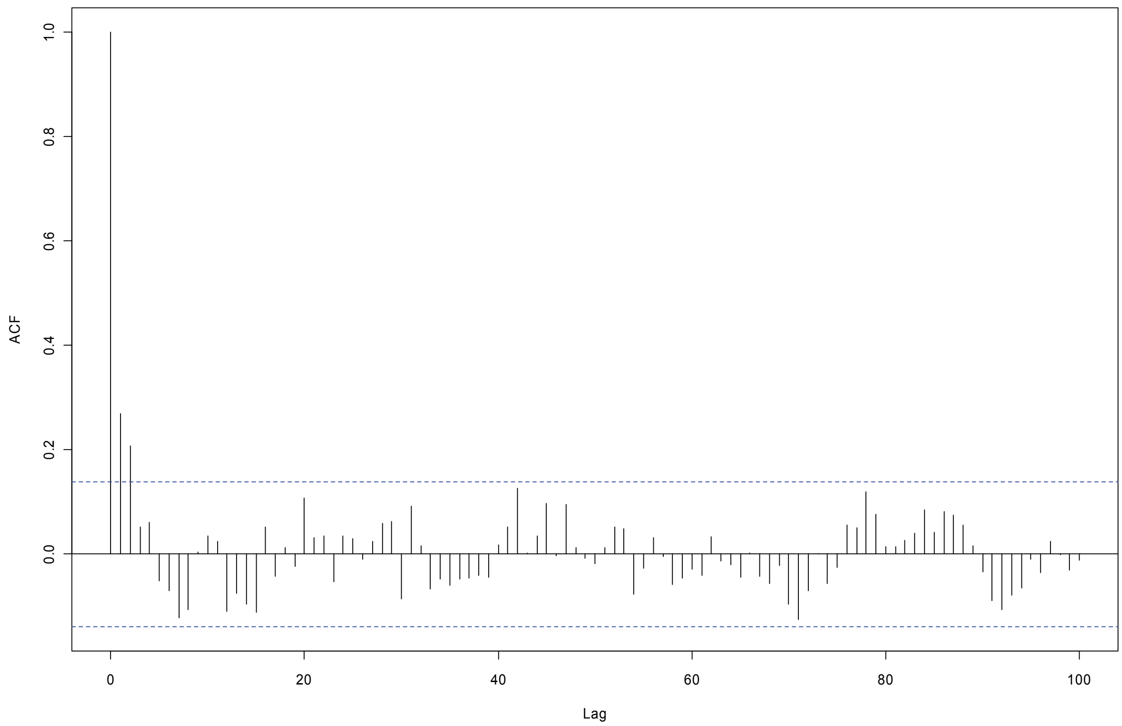


Figure 8: Spectral Densities of Residual Series of Figure 4

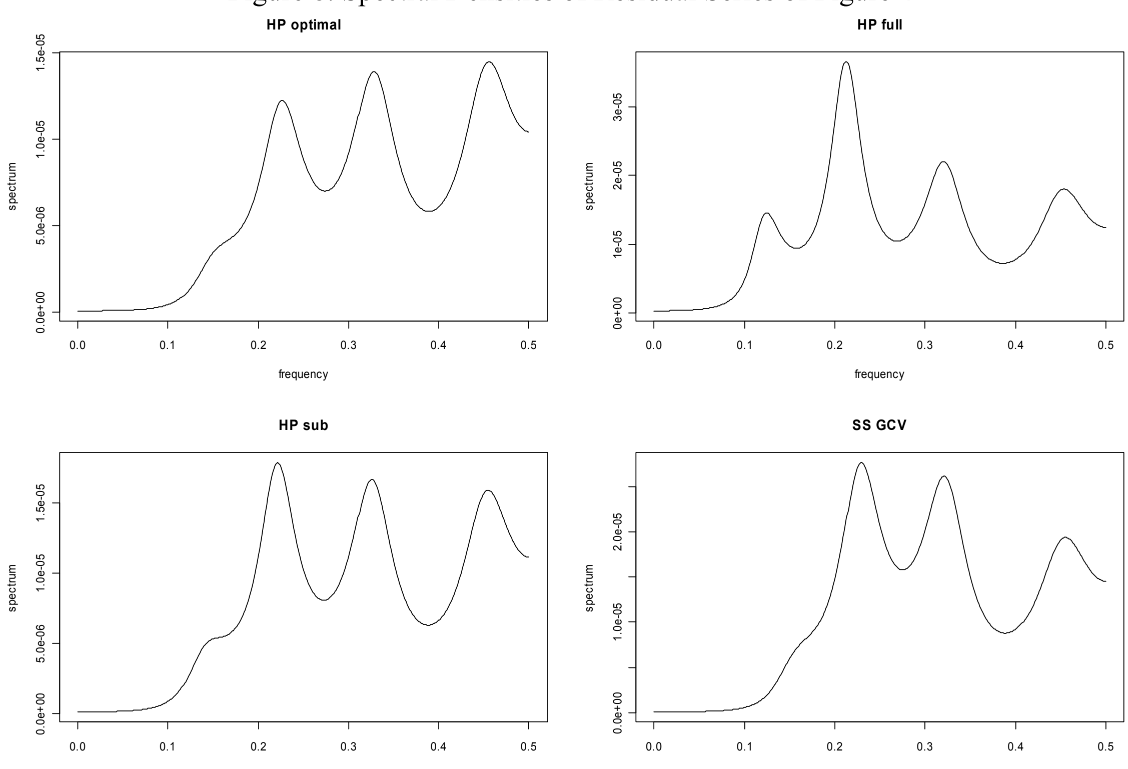


Figure 9: Spectral Densities of Residual Series of Figure 5

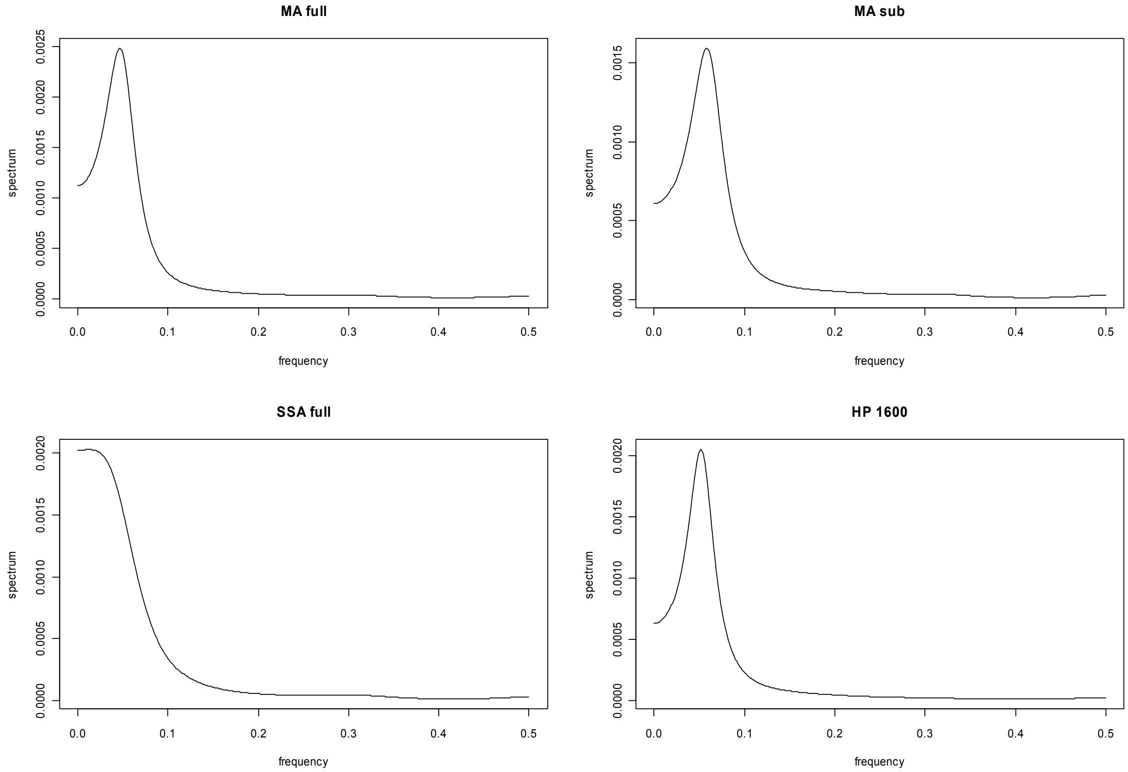
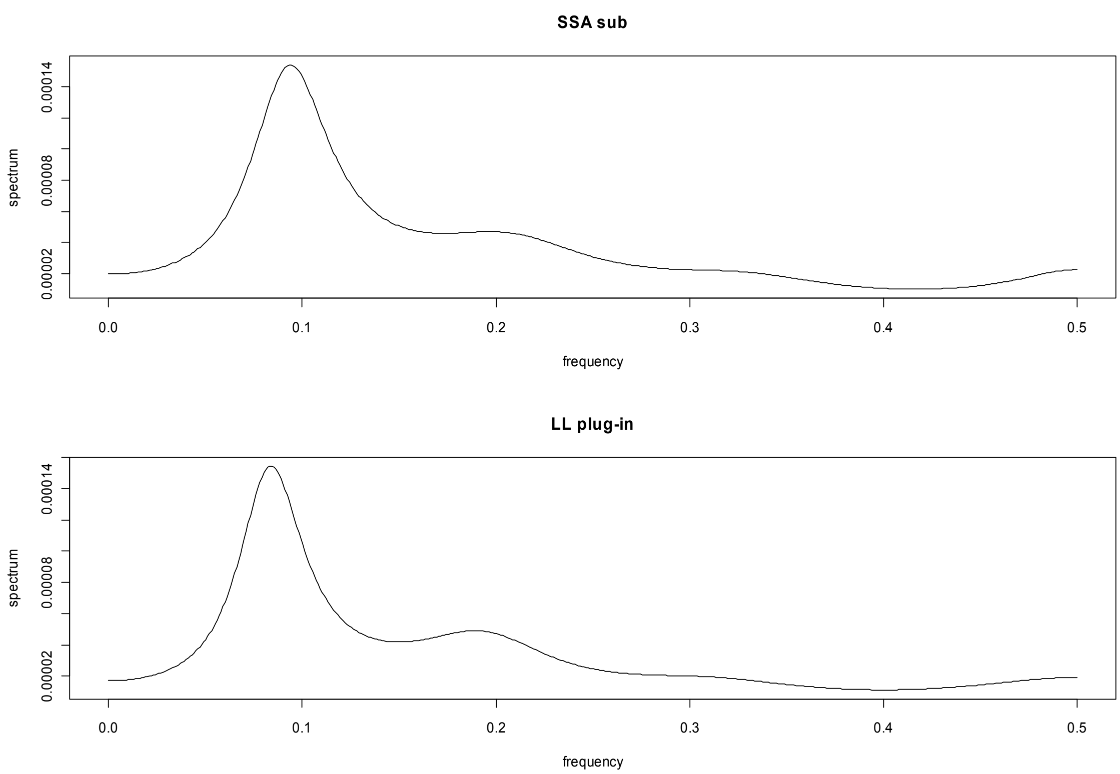


Figure 10: Spectral Densities of Residual Series of Figure 6



the local linear smoother with the plug-in bandwidth appear to be a reasonable, economically viable compromise as those smoothers that capture the essence of the trend in U.S. output.

Based on the above discussion findings from this study may be summarized as follows:

- The proposed methodology can be used to achieve the same degree of smoothing for the HP smoother as that implied by other, more sophisticated, approaches.
- A number of alternative smoothers lead to the same shape and properties for the output gap as the HP smoother and these smoothers can be clustered together based on the shape of the series and their corresponding spectral densities.

Combining MA or SSA smoothers with subsampling leads to essentially the same results as the ones obtained by the default HP smoother.

Analyses herein illustrate a high potential for the application of less sophisticated, universally applicable, smoothing methods in trend extraction. This article proposes a simple, intuitive and immediately applicable method for selecting the degree of smoothing for such methods. One of the advantages of the having such methods available is that they can be used for benchmarks against which other, more sophisticated methods, can be compared.

Conclusion

This article proposed a new methodology for selecting the degree of smoothing in problems of trend extraction. The method uses an alternative to, mean squared error, proximity criterion which is minimized for selecting the required value of the smoothing parameter. This criterion is based on the average sign of the residuals obtained after smoothing and its minimization implies a probabilistic symmetrization of the residuals: it was shown that the minimizing value implies that the resulting residuals have a zero median.

The viability and usefulness of the proposed method is illustrated using simulations where the underlying type of trend is known. The results from these simulations are suggestive that the method is competitive because it can perform on par with - or better than - existing methods. In particular, it was shown that less sophisticated smoothing methods, like the moving average, for which no formal method for selecting the degree of smoothing exist, can be made to perform on par with more sophisticated methods. The use of sub-sampling can also help in improving performance.

A number of extensions can be undertaken based on the current work include the following:

- Consider the construction of confidence bands around the trend; since the method of this paper results in residuals with zero median such confidence bands can be based on the quantiles of the residual distribution.
- Consider a more systematic, expanded comparison between smoothing methods and approaches for selecting the optimal degree of smoothing.
- Apply the method of this article in the context of non-parametric autoregressive models and examine whether it can successfully be used in selecting both the degree of smoothing and the order of the model.

References

Christiano, L., & Fitzgerald, T. (2003). The band pass filter. *International Economic Review*, *44*, 435-465.

Dermoune, A., Djehiche, B., & Rahmania, N. (2007). A consistent estimator of the smoothing parameter in the Hodrick-Prescott filter. *Journal of the Japan Statistical Society*, forthcoming.

Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Monographs on Statistics and Applied Probability, Chapman and Hall.

Fan, J., & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. Monographs on Statistics and Applied Probability, Chapman and Hall.

Härdle, W. (1990). *Applied nonparametric regression*. Boston, MA: Cambridge University Press.

Hart, J. D. (1997). *Nonparametric smoothing and lack of fit tests*. New York: Springer.

Hodrick, R., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, *29*, 1-16.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303-1313.

Thomakos, D. D. (2008). Optimal linear filtering, smoothing and trend extraction for processes with unit roots and cointegration. *Working paper #24, Department of Economics, University of Peloponnese*.