A CONSISTENT ESTIMATOR OF THE SMOOTHING PARAMETER IN THE HODRICK-PRESCOTT FILTER

Azzouz Dermoune, * Boualem Djehiche, ** and Nadji Rahmania ***

The so-called Hodrick-Prescott filter was first introduced in actuarial science to estimate trends from claims data and now is widely used in economics and finance to estimate and predict e.g. business cycles and trends in financial data series. This filter depends on the noise-to-signal ratio \( \alpha \) that acts as a smoothing parameter. We propose a new consistent estimator of this smoothing parameter and construct corresponding non-asymptotic confidence intervals with a precise confidence level.

**Key words and phrases:** Adaptive estimation, Hodrick-Prescott filter, orthogonal parametrization, Gaussian process.

1. Introduction

Leser (1961) and later on Hodrick and Prescott (1980) and (1997) defined a trend \( y = (y_1, \ldots, y_T) \) of a time series \( x = (x_1, \ldots, x_T) \) as the minimizer of \( \sum_{t=1}^{T} (x_t - y_t)^2 + \alpha \sum_{t=1}^{T-2} (y_{t+2} - 2y_{t+1} + y_t)^2 \) for an appropriately chosen positive parameter \( \alpha \), called the smoothing parameter. Assuming a Gaussian random walk model of the trend, Hodrick and Prescott (1997) suggest that \( \alpha \) should be the ratio of the variance of noise to the variance of the signal. When the time series \( x \) represents the log of U.S. real GNP, using macroeconomic arguments, they find that \( \alpha = 1600 \), is ought to be optimal to get a smooth trend close to the actual real GNP. This procedure for filtering a trend as a smooth curve from the data has a long history in actuarial science starting with the graduation method developed by Whittaker (1923) and Henderson (1924) that was the starting point for the actual filter, first proposed by Leser (1961) and later on introduced into economics by Hodrick and Prescott (1980). See e.g. Stigler (1978) and Green and Silverman (1994) for an overview.

There are two different approaches (see e.g. Green and Silverman (1994)) to the question of choosing the noise-to-signal ratio \( \alpha \). The first approach is to consider the free choice of \( \alpha \) as an advantage. The features of the data can be explored by varying it. To this approach belongs the Bayesian framework, discussed in Trimbure (2006), that allows for the influence of non-sample information on the choice of \( \alpha \). This is done by further specifying prior distributions of \( \alpha \) in the underlying unobserved components model. This helps ensure an appropriate degree of smoothness in the estimated trend while allowing for uncertainty in \( \alpha \). The other view advocates for a need of an automatic estimation method based on the observed sample data (sample information). Generalized cross-validation techniques are among the automatic methods used to estimate the smoothing parameter \( \alpha \) (see Craven and Wahba (1978) and Green and Silverman (1994) for further

---

*Laboratoire de Probabilités et Statistique, UFR de Mathématiques, USTL, Bât. M2, 59655 Villeneuve d’Ascq Cédex, France. e-mail: azzouz.dermoune@univ-lille1.fr

**Department of Mathematics, The Royal Institute of Technology, S-100 44 Stockholm, Sweden. e-mail: boualem@math.kth.se

***Laboratoire de Probabilités et Statistique, UFR de Mathématiques, USTL, Bât. M2, 59655 Villeneuve d’Ascq Cédex, France. e-mail: Nadji.Rahmania@univ-lille1.fr
Schlicht (2005) proposes another automatic method, where the estimator of the noise-to-signal ratio is a fixed point of a highly nonlinear equation, obtained using the maximum-likelihood and moments matching techniques. The consistency of Schlicht’s estimator is however still an open problem.

The main result of this paper is to suggest a new estimator of the smoothing parameter \( \alpha \) which is explicit and consistent, based on the optimality criterium suggested by Schlicht (2005, Theorem 1). We also construct a confidence interval with a precise confidence level.

The choice of the appropriate approach to determine the noise-to-signal ratio \( \alpha \), relies very much on the amount of information we want to be reflected by the trend or, alternatively, by the residual term. In some situations e.g. stock and currency data, many asset allocation strategies are based on a correct analysis of the trend, in which case, it is optimal that the residual term is a pure white noise. Schlicht’s optimality criterium is designed to provide such a feature of the residual terms. In other situations, the preferred choice of \( \alpha \) is the one that makes the residual term cyclical, although not necessarily white noise, as seems to be the case for macroeconomic time series. As it is seen from the autocorrelation functions displayed in Figure 2, in Section 4, below, the residual term corresponding to the Hodrick-Prescott choice of the parameter value \( \alpha = 1600 \), applied to the US real GNP time series, generates a residual term that is cyclical but it is not white noise, as assumed by the model. But, the residual term corresponding to the value \( \alpha = 2.40 \) obtained by the Schlicht’s optimality criterium or our statistics (\( \alpha = 0.41 \)) displays a closer behaviour to white noise.

The paper is organized as follows. In Section 2, we introduce the Hodrick-Prescott filter and in Section 3, we derive a consistent estimator of the smoothing parameter \( \alpha \), and non-asymptotic confidence intervals, given a confidence level, when the trend is a Gaussian random walk, based on the optimality criterium suggested by Schlicht (2005, Theorem 1). Finally, in Section 4, we apply our estimator to extract the trend associated with a set of simulated data and real market data including the US real GNP, the spot prices of the major currencies and stock indexes. The quite good fit of the trend to the considered time series, suggests that this approach to trend estimation may be useful for practitioners, due to the relatively simple form of the estimator.

2. The Hodrick-Prescott filter

Let \( x = (x_1, \ldots, x_T) \in \mathbb{R}^T \) be a time series of observables. The Hodrick-Prescott filter (HP in short) decomposes \( x \) into a nonstationary trend \( y \in \mathbb{R}^T \) and a cyclical residual component (noise term) \( u \in \mathbb{R}^T \):

\[
(2.1) \quad x = y + u.
\]

Given a smoothing parameter \( \alpha > 0 \), this decomposition of \( x \) is obtained by minimizing the weighted sum of squares

\[
(2.2) \quad \|x - y\|^2 + \alpha \|D^2 y\|^2
\]

with respect to \( y \), where for \( a \in \mathbb{R}^T \), \( \|a\|^2 = \sum_{t=1}^{T} a_t^2 \).

Here, \( D^2 y \) is the trend disturbance obtained by acting the second order forward shift operator \( D^2 \) on the trend \( y = (y_1, y_2, \ldots, y_T) \):

\[
D^2 y_t := (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t), \quad t = 1, 2, \ldots, T - 2,
\]

or, equivalently,

\[
D^2 y_t := 2 \left( \frac{y_{t+2} + y_{t+1}}{2} - y_{t+1} \right), \quad t = 1, 2, \ldots, T - 2,
\]
measuring the deviation between the value of the trend at \( t + 1 \), \( y_{t+1} \) and the linear interpolation between \( y_t \) and \( y_{t+2} \).

In vector form,

\[
P y(t) = D^2 y_t, \quad t = 1, \ldots, T - 2,
\]

where, the shift operator \( P \) is the following \((T - 2) \times T\)-matrix

\[
P := \begin{pmatrix}
1 & -2 & 1 & \cdots & 0 \\
0 & 1 & -2 & 1 & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 1 & -2 & 1
\end{pmatrix}.
\]

As pointed out e.g. in Pedersen (2001), the first term in (2.2) measures a goodness-of-fit by minimizing the deviation between the trend \( y_t \) and the observation \( x_t \) and the second term is a measure of the degree-of-smoothness which penalizes decelerations in growth rate of the trend component, by minimizing the deviation between the trend value \( y_{t+1} \) and the linear interpolation between \( y_t \) and \( y_{t+2} \).

Both Reeves et al. (2000) and Araujo et al. (2003), found the third order shift operator more appropriate than the second order one to extract the trend from some Foreign Exchange Rate series. The matrix formulation to the HP-filter adopted in this paper fully extends to higher order shift operators or any other smoothing operator.

Since \( P \) is of rank \( T - 2 \), the signal \( v := Py \) does not determine a unique \( y \) but rather the set of solutions (see Schlicht (2005) for further details)

\[
y := \{P'(PP')^{-1}v + Z\gamma; \quad \gamma \in \mathbb{R}^2\}
\]

where the \( T \times 2 \)-matrix \( Z \) satisfies

\[
PZ = 0, \quad Z'Z = I_2,
\]

with \( I_2 \) denoting the \( 2 \times 2 \) identity matrix. In view of Eq. (2.1), the time series \( x \) can be represented in terms of \((u, v)\) as

\[
x = u + P'(PP')^{-1}v + Z\gamma,
\]

for some \( \gamma \in \mathbb{R}^2 \).

As the matrix \((I_T + \alpha P'P)\) is positive definite, the unique solution \( y(\alpha, x) \) to the optimal problem (2.2) is

\[
y(\alpha, x) = (I_T + \alpha P'P)^{-1}x,
\]

where \( I_T \) denotes the \( T \times T \) identity matrix. Eq. (2.6) defines the descriptive filter that associates a trend \( y \) to the time series \( x \), depending on the smoothing parameter \( \alpha \) and the disturbance operator \( P \).

Following Schlicht (2005), a way to estimate the smoothing parameter \( \alpha \) is to let the optimal solution \( y(\alpha, x) \) in Eq. (2.6) be the best predictor of any trend \( y \) given the time series \( x \), i.e.

\[
y(\alpha, x) = E[y \mid x].
\]

This approach of estimating \( \alpha \) assumes that we are able to compute explicitly this conditional expectation, which is not always the case. The normal and more generally the elliptical probability distributions are among the few models for which an explicit formula for the conditional expectation is possible. In order to estimate the trend and the smoothing parameter, given the time series of observations \( x \), we obviously need a model for the joint distribution of \((x, y)\). Using (2.3) and (2.5), this can be achieved through imposing a model for the joint distribution of \((u, v)\).
3. A Gaussian random walk model of the trend

In the literature (cf. e.g. Hodrick and Prescott (1997) and Schlicht (2005)), a widely used model (and perhaps the only feasible case) for the joint distribution of \((u, v)\), is to assume that the disturbances \(u\) and \(v\) independent and normally distributed. This turns \((x, y)\) into a normally distributed vector, which makes the estimation issue of the trend \(y\) and the smoothing parameter \(\alpha\), using (2.6) and (2.7), feasible.

3.1. Independent and identically distributed disturbances

In particular, assuming furthermore that the noise term \(u\) and and the signal term \(v\) have zero means and covariance matrices \(\sigma_u^2 I_T\) and \(\sigma_v^2 I_{T-2}\), where \(I_T\) and \(I_{T-2}\) denote the \(T \times T\) and \((T-2) \times (T-2)\) identity matrices, respectively:

\[
\begin{pmatrix}
u \\
v
\end{pmatrix} \sim N(0, \Sigma_{uv}),
\]

with covariance matrix

\[
\Sigma_{uv} := \begin{pmatrix}
\sigma_u^2 I_T & 0 \\
0 & \sigma_v^2 I_{T-2}
\end{pmatrix},
\]

makes the increments of the trend \(y\) following a Gaussian random walk, since, by Eq. (2.3), \(y_{t+2} - y_{t+1} = y_{t+1} - y_t + v_t\). This turns the time series \(x\) into a trend \(y\) generated by a Gaussian random walk and a normal disturbance \(u\). That is, in view of (2.5), \((x, y)\) is normally distributed:

\[
\begin{pmatrix}
x \\
y
\end{pmatrix} \sim N\left(\begin{pmatrix}
Z \\
Z
\end{pmatrix} \gamma, \Sigma_{xy}\right),
\]

with covariance matrix

\[
\Sigma_{xy} := \begin{pmatrix}
\sigma_u^2 I_T + \sigma_v^2 Q & \sigma_v^2 Q \\
\sigma_v^2 Q & \sigma_v^2 I_{T-2}
\end{pmatrix},
\]

where,

\[
Q := P'(PP')^{-1}(PP')^{-1}P.
\]

This yields an explicit expression of the conditional expectation of the trend \(y\) given the time series \(x\):

\[
E[y|x] = Z\gamma + \sigma_v^2 Q \left[\sigma_u^2 I_T + \sigma_v^2 Q\right]^{-1} (x - Z\gamma).
\]

Now, in view of (2.7), the smoothing parameter \(\alpha\) and the parameter \(\gamma\) satisfy

\[
Z\gamma + \sigma_v^2 Q \left[\sigma_u^2 I_T + \sigma_v^2 Q\right]^{-1} (x - Z\gamma) = (I_T + \alpha P'P)^{-1} x.
\]

In the following proposition, we recall Theorem 1 in Schlicht (2005) that gives a characterization of the optimal smoothing parameter \(\alpha\) as the noise-to-signal ratio, \(\alpha = \sigma_u^2/\sigma_v^2\), and the free parameter \(\gamma\) that satisfy Eq. (3.4) in terms of the time series of observations \(x\), the second order forward shift operator \(P\) and \(Z\).

**Proposition 1.** (Schlicht (2005), Theorem 1)

(a) If Equation (3.4) holds then \(\gamma = Z'x\).

(b) Equation (3.4) holds if and only if \(\gamma = Z'x\) and \(\alpha = \sigma_u^2/\sigma_v^2\).
3.2. A consistent estimator of the noise-to-signal ratio $\alpha$

In this section we propose explicit unbiased consistent estimators of the variances $\sigma_u^2$ and $\sigma_v^2$, in the sense that, as the length $T$ of the series of observations tends to infinity, the proposed estimators converge in probability, with normally distributed errors to the respective variances. A consistent estimator of the smoothing parameter $\alpha$ will be the ratio of the estimators of the respective variances of $u$ and $v$.

To this end, we use observations from the centered time series $P\mathbf{x}$:

$$P\mathbf{x} = v + Pu \sim \mathcal{N}(0, \sigma_u^2 I_{T-2} + \sigma_v^2 PP').$$

First, we note that the Gaussian time series $P\mathbf{x}$ is stationary. In fact the elements $V(i, j)$ of the $(T-2 \times T-2)$ covariance matrix of $P\mathbf{x}$ satisfy

$$V(i, j) = \sigma_v^2 \delta_{ij} + \sigma_u^2 (PP')_{ij} = r_{|i-j|},$$

where,

$$r_k = \begin{cases} 
\sigma_v^2 + 6\sigma_u^2, & \text{if } k = 0; \\
-4\sigma_u^2, & \text{if } k = 1; \\
\sigma_u^2, & \text{if } k = 2; \\
0, & \text{otherwise.}
\end{cases}$$

Now, it is a classical fact that (see e.g. Chapter 7 in Brockwell and Davis (1991) or Proposition 2.1 in Giurcanu and Spokoiny (2002))

$$\hat{r}_k = \frac{1}{(T-2) - k} \sum_{j=1}^{T-2-k} P\mathbf{x}(j) P\mathbf{x}(j + k), \quad k = 0, 1, 2,$$

is an unbiased estimator of $r_k = E[P\mathbf{x}(s) P\mathbf{x}(s + k)]$ in the sense that $E[\hat{r}_k] = r_k$.

We want an explicit form of the quadratic error

$$a_k^2 := \text{Var}(\hat{r}_k) \quad k = 0, 1, 2.$$

To this purpose, let $A_k$ be the $(T-2) \times (T-2)$ matrix with the entries

$$\frac{1}{2(T-2-k)} (\delta(s - t - k) + \delta(t - s - k)).$$

We recall the following result due to Giurcanu and Spokoiny (2002).

**Proposition 2.** We have, for $k = 0, 1, 2$,

$$\hat{r}_k = (P\mathbf{x})^t A_k P\mathbf{x}, \quad r_k = tr(A_k V)$$

and

$$a_k^2 = 2tr(A_k V)^2.$$

By a simple computation we have for $T \geq 4$,

$$a_0^2 = \frac{1}{(T-2)^2} [2(T-2)r_0^2 + 4(T-3) r_1^2 + 4(T-4) r_2^2],$$

(3.6)

$$a_1^2 = \frac{1}{(T-3)^2} [(T-3)(r_1^2 + r_0^2) + 2(T-4)(r_1 r_0 + r_2^2) + 2(T-5) r_3^2]$$

(3.7)
\[ a_2^2 = \frac{1}{(T-4)^2}[(T-4)(r_0^2 + r_0^n) + 2(T-5)r_1^2 + 2(T-6)r_2^2]. \]

From that we derive the following explicit central limit for the estimators \( \hat{r}_k \), \( k = 0, 1, 2 \).

**Corollary 1.**

1. \( \sqrt{T}[\hat{r}_0 - r_0] \to N(0, 2r_0^2 + 4r_1^2 + 4r_2^2). \)
2. \( \sqrt{T}[\hat{r}_1 - r_1] \to N(0, (r_0 + r_2)^2 + r_2^2 + 3r_1^2). \)
3. \( \sqrt{T}[\hat{r}_2 - r_2] \to N(0, r_0^2 + 2r_1^2 + 3r_2^2). \)

The easily checked relation

\[ E[\hat{r}_1] = -4\sigma_u^2 \]

suggests the following consistent unbiased estimator of \( \sigma_u^2 \):

\[ \hat{\sigma}_u^2 = \frac{1}{4}\hat{r}_1 = -\frac{1}{4(T-3)} \sum_{j=1}^{T-3} Px(j)Px(j+1). \]

Furthermore, since

\[ E[\hat{r}_0] = \sigma_v^2 + 6\sigma_u^2, \]

in view of (3.9), we get a consistent unbiased estimator of \( \sigma_v^2 \):

\[ \hat{\sigma}_v^2 = \hat{r}_0 + \frac{3}{2}\hat{r}_1, \]

or

\[ \hat{\sigma}_v^2 = \frac{1}{T-2} \sum_{j=1}^{T-2} Px(j)^2 + \frac{3}{2(T-3)} \sum_{j=1}^{T-3} Px(j)Px(j+1). \]

Therefore, in view of Proposition 1, we get a consistent estimator of the smoothing parameter \( \alpha \):

**Theorem 1.** The following statistics

\[ \hat{\alpha} = \frac{1}{4} \left( \frac{3}{2} + \frac{(T-3) \sum_{j=1}^{T-3} Px(j)^2}{(T-2) \sum_{j=1}^{T-3} Px(j)Px(j+1)} \right)^{-1} \]

based on the time series of observation \( Px \), is a consistent estimator of the smoothing parameter \( \alpha \).

### 3.3. Non-asymptotic confidence intervals for the noise-to-signal ratio \( \alpha \)

In this section we derive non-asymptotic confidence intervals for \( \alpha \). From the calculations in Section 3.2, it turns out that the parameter

\[ \beta := \frac{1 + 6\alpha}{\alpha} \]

and its estimator

\[ \hat{\beta} := \frac{1 + 6\hat{\alpha}}{\hat{\alpha}} = -4\frac{\hat{r}_0}{\hat{r}_1}. \]
are more plausible to study than \( \alpha \) and \( \hat{\alpha} \). Therefore, we will derive non-asymptotic confidence intervals for \( \beta \) in terms of \( \hat{\beta} \). We first suggest non-asymptotic confidence intervals for \( r_0 \) and \( r_1 \) in terms of their unbiased estimators \( \hat{r}_0 \) and \( \hat{r}_1 \) and \( \beta \).

We have the following tail estimates of \( \hat{r}_k - r_k \), \( k = 0, 1 \) and \( a_0 \) and \( a_1 \) defined in Eqs. (3.6) and (3.7).

**Proposition 3.**

(1) If \( T > 2 \) and \( 0 < \lambda \leq \frac{\sqrt{8(T-2)}}{3} \), then
\[
P(|\hat{r}_0 - r_0| \geq a_0 \lambda) \leq \exp(-\frac{\lambda^2}{4}).
\]

(2) If \( T > 3 \) and \( 0 < \lambda \leq \frac{\sqrt{(T-3)}}{3} \), then
\[
P(|\hat{r}_1 - r_1| \geq a_1 \lambda) \leq \exp(-\frac{\lambda^2}{4}).
\]

(3) If \( T > 3 \) and \( 2 \log 2 < \lambda \leq \frac{\sqrt{(T-3)}}{3} \), then
\[
P(\max_{k=0,1} |\hat{r}_k - r_k| \geq \frac{r_0 \lambda \sqrt{10}}{\sqrt{T - 3}}) \leq 2 \exp(-\frac{\lambda^2}{4}).
\]

**Proof.** See the appendix below.

In the next proposition we give non-asymptotic confidence intervals (which depend on \( \beta \)) for \( r_0 \) and \( r_1 \). Denote
\[
a := \frac{\lambda \sqrt{10}}{4\sqrt{T - 3}}
\]

**Proposition 4.** Assume that \( T > 3 \) and \( 2 \sqrt{\log 2} < \lambda < \sqrt{\frac{T-3}{10}} \), then with probability at least \( 1 - 2 \exp(-\frac{\lambda^2}{4}) \)

\[
r_0 \in \left[ \frac{\hat{r}_0}{1 + 4a}, \frac{\hat{r}_0}{1 - 4a} \right]
\]

and, when \( \beta < 1/a \),

\[
r_1 \in \left[ \frac{\hat{r}_1}{1 - a\beta}, \frac{\hat{r}_1}{1 + a\beta} \right],
\]

whereas, when \( \beta > 1/a \),

\[
r_1 \in \left( -\infty, \frac{\hat{r}_1}{1 + a\beta} \right].
\]

**Proof.** Recall that \( r_0 = -\frac{\beta}{4} r_1 \). For \( T > 3 \) and \( 2 \sqrt{\log 2} < \lambda < \sqrt{\frac{T-3}{10}} \) we get
\[
2 \sqrt{\frac{(T-3) \log 2}{10}} < \lambda < 1/4.
\]

Proposition 3.5 yields that
\[
P \left( |r_0 - \hat{r}_0| \leq 4ar_0, |r_1 - \hat{r}_1| \leq 4ar_0 \right) \geq 1 - 2 \exp(-\frac{\lambda^2}{4}),
\]
from which it follows that with probability at least $1 - 2\exp(-\frac{\lambda^2}{4})$

$$r_0 \in \left[\frac{\hat{r}_0}{1 + 4a}, \frac{\hat{r}_0}{1 - 4a}\right],$$

and, when $\beta < 1/a$,

$$r_1 \in \left[\frac{\hat{r}_1}{1 - a\beta}, \frac{\hat{r}_1}{1 + a\beta}\right],$$

whereas, when $\beta > 1/a$,

$$r_1 \in \left(-\infty, \frac{\hat{r}_1}{1 + a\beta}\right].$$

We are now ready to give non-asymptotic confidence intervals for $\beta$. Recall $\beta$, $\hat{\beta}$ and $a$ from (3.12), (3.13) and (3.14).

**Theorem 2.** Assume that $T > 3$ and $2\sqrt{\log 2} < \lambda < \sqrt{\frac{T - \beta}{10}}$. Then,

(1) When $\beta < 1/a$,

$$P \left( \beta \in \left[\hat{\beta}_1 1_A, \hat{\beta}_2 1_A\right] \right) \geq 1 - 2\exp(-\frac{\lambda^2}{4}) - P(A^c).$$

where

$$A := A(\lambda, T) = \left\{ \omega; \hat{\beta}(\omega) < \frac{1 - 4a}{a} \right\},$$

$$\hat{\beta}_1 = \frac{\hat{\beta}}{1 + (4 + \hat{\beta})a},$$

$$\hat{\beta}_2 = \frac{\hat{\beta}}{1 - (4 + \hat{\beta})a}.$$

Here, $1_A$ denotes the indicator function of the set $A$.

(2) When $\beta > 1/a$,

$$P \left( \beta \in \left(0, \hat{\beta}_3 1_B\right] \right) \geq 1 - 2\exp(-\frac{\lambda^2}{4}) - P(B^c).$$

where

$$B := B(\lambda, T) = \left\{ \omega; \hat{\beta}(\omega) < \frac{1}{a} \right\},$$

$$\hat{\beta}_3 = \frac{\hat{\beta}}{1 - a\beta}.$$

and, $1_B$ denotes the indicator function of the set $B$.

**Proof.** We only consider the case $\beta < 1/a$. The case $\beta > 1/a$ is treated in a similar fashion.

By Proposition 4, it follows from (3.15) and (3.16) that

$$\frac{\hat{r}_0}{1 + 4a} \leq r_0 \leq \frac{\hat{r}_0}{1 - 4a}.$$
and
\[ \frac{\hat{r}_1}{1 + a\beta} \leq r_1 \leq \frac{\hat{r}_1}{1 - a\beta} \]
with probability at least \(1 - 2 \exp(-\frac{\lambda^2}{4})\). Therefore, using \(\beta = -4\frac{r_0}{r_1}\), we get
\[
P \left( \frac{\hat{\beta}(1 - a\beta)}{1 + 4a} \leq \beta \leq \frac{\hat{\beta}(1 + a\beta)}{1 - 4a} : A \right) + P(A^c) \geq 1 - 2 \exp(-\frac{\lambda^2}{4}).
\]
Consider the functions
\[
f(\beta) = \left(1 + \frac{\hat{\beta}a}{1 + 4a}\right) \beta - \frac{\hat{\beta}}{1 + 4a}
\]
and
\[
g(\beta) = \left(1 - \frac{\hat{\beta}a}{1 - 4a}\right) \beta - \frac{\hat{\beta}}{1 - 4a}.
\]
and let \(\hat{\beta}_1\) and \(\hat{\beta}_2\) denote their respective zeros, where
\[
\hat{\beta}_1 = \frac{\hat{\beta}}{1 + (4 + \hat{\beta})a}
\]
and
\[
\hat{\beta}_2 = \frac{\hat{\beta}}{1 - (4 + \hat{\beta})a}.
\]
Since, \(\lambda < \sqrt{\frac{T - 3}{10}}\) that is \(a < 1/4\), both the slope of the function \(g\) is positive and \(0 < \hat{\beta}_1 < \hat{\beta}_2\) on the set \(A\).
Therefore,
\[
(3.25) \quad P \left( \beta \in \left[ \hat{\beta}_1 1_A, \hat{\beta}_2 1_A \right] \right) \geq 1 - 2 \exp(-\frac{\lambda^2}{4}) - P(A^c).
\]
Remark 1.

(1) Since, by (3.14), \(a \to 0\) as the time horizon \(T \to \infty\), it is only the case \(\beta < 1/a\) that is of importance. Moreover, by Proposition 3, the consistency of the estimator \(\hat{\beta}\) yields
\[
\lim_{T \to \infty} P(A^c) = 0.
\]
Hence, asymptotically, \(\beta\) belongs to \(\left[ \hat{\beta}_1, \hat{\beta}_2 \right]\) with probability at least \(1 - 2 \exp(-\frac{\lambda^2}{4})\), provided that \(\lambda > 2\sqrt{\log 2}\).

(2) In terms of \(\hat{\alpha}\), the bounds (3.24) and (3.24) read
\[
(3.26) \quad \hat{\alpha}_1 = \frac{1}{\hat{\beta}_2 - 6} = \frac{(1 - 10a)\hat{\alpha} - a}{1 + 6a(1 + 10\hat{\alpha})}
\]
and
\[
(3.27) \quad \hat{\alpha}_2 = \frac{1}{\hat{\beta}_1 - 6} = \frac{(1 + 10a)\hat{\alpha} + a}{1 - 6a(1 + 10\hat{\alpha})},
\]
with \( \hat{\alpha}_1 < \hat{\alpha}_2 \) on the set \( A \). Therefore,

\[
(3.28) \quad \mathbf{P} \left( \alpha \in \left[ \hat{\alpha}_1 \mathbf{1}_A, \hat{\alpha}_2 \mathbf{1}_A \right] \right) \geq 1 - 2 \exp\left( -\frac{\lambda^2}{4} \right) - \mathbf{P}(\mathbf{A}^c).
\]

Note that, since \( \beta = \frac{1}{\alpha} + 6 > 6 \), the confidence intervals \( [\hat{\beta}_1, \hat{\beta}_2] \) should suitably satisfy \( 6 < \hat{\beta}_1 < \hat{\beta}_2 \).

4. Simulation and calibration

To illustrate the performance of the estimator \( \hat{\alpha} \), we give a numerical example based on large simulations drawn from the distribution of \( P \mathbf{x} \) for different values of the time horizon \( T \) and specifying fixed values of \( \sigma_u^2 \) and \( \sigma_v^2 \). Moreover, we construct a trend from market data including the quarterly US real GNP and monthly spot prices of the major currencies and stock indexes.

4.1. A numerical example

In this section we report for the performance of different estimates of the statistic \( \hat{\alpha} \) based on simulations drawn from the distribution of \( P \mathbf{x} \):

\[
P \mathbf{x} \sim \mathcal{N}(0, \sigma^2_{P \mathbf{x}}),
\]

for different values of \( T - 2 \) (the dimension of the matrix \( P \mathbf{P}' \)), given the true value of the smoothing parameter is \( \alpha \). At this stage, we consider two cases. The case where \( \alpha \leq 1 \), which corresponds to \( \sigma_u \leq \sigma_v \), meaning that the trend \( y \) explains the variation of the time series \( x \) more than the residual noise \( u \). As we will see in the numerical examples below, in this case our estimator \( \hat{\alpha} \) exhibits smaller variances for larger values of \( T \), suggesting its suitably for estimation \( \alpha \).

However, when \( \alpha > 1 \) i.e. \( \sigma_u > \sigma_v \), meaning that the residual noise term \( u \) explains the variation of the time series \( x \) more than the trend, our estimator \( \hat{\alpha} \) exhibits larger variances for large values of \( T \), suggesting that it may not be good enough to estimate larger values of \( \alpha \). This is better explained through the parameter \( \beta \). Indeed, by (3.12), we have

\[
\alpha = \frac{1}{\beta - 6},
\]

which has a singularity at \( \beta = 6 \). Obviously, \( \alpha > 1 \) corresponds to \( 6 < \beta < 7 \). The closer \( \beta \) is to 6, the larger is \( \alpha \). From the relation

\[
\hat{\alpha} = \frac{1}{\hat{\beta} - 6},
\]

we should expect that (see Table 5, below), when \( \beta \) is close to 6, the goodness of \( \hat{\beta} \) (small variance), as estimator of \( \beta \), implies larger variance of \( \hat{\alpha} \).
• **The case** $\alpha \leq 1$. In Tables 1 to 4, we report for the performance of different estimates of the statistic $\hat{\alpha}$ based on 1000 respectively 5000 simulations drawn from the distribution of $Px$ for the values of $T - 2$ (the dimension of the matrix $PP'$) ranging from 500 to 5000 and assuming that $\sigma_u^2 = \sigma_v^2 = 1$, respectively $\sigma_u^2 = 1$, $\sigma_v^2 = 2$ i.e. the true value of the smoothing parameter is $\alpha = 1$, respectively, $\alpha = 0.5$. The reported figures in Tables 1 and 2, show that the statistic $\hat{\alpha}$ converges to the true values $\alpha = 1$ resp. $\alpha = 0.5$ with decreasing standard errors, as the time horizon $T$ gets large.

Table 1. Performance of different estimates of the statistics $\hat{\alpha}$ of $\alpha = 1$ based on 1000 simulations.

<table>
<thead>
<tr>
<th>T-2</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\hat{\alpha}$)</td>
<td>1.13</td>
<td>1.05</td>
<td>1.03</td>
<td>1.016</td>
<td>1.005</td>
</tr>
<tr>
<td>standard deviation($\hat{\alpha}$)</td>
<td>0.54</td>
<td>0.29</td>
<td>0.25</td>
<td>0.17</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 2. Performance of different estimates of the statistics $\hat{\alpha}$ of $\alpha = 0.5$ based on 1000 simulations.

<table>
<thead>
<tr>
<th>T-2</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\hat{\alpha}$)</td>
<td>0.514</td>
<td>0.506</td>
<td>0.506</td>
<td>0.504</td>
<td>0.502</td>
</tr>
<tr>
<td>standard deviation($\hat{\alpha}$)</td>
<td>0.13</td>
<td>0.09</td>
<td>0.07</td>
<td>0.005</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 3. Performance of different estimates of the statistics $\hat{\alpha}$ of $\alpha = 1$ based on 5000 simulations.

<table>
<thead>
<tr>
<th>T-2</th>
<th>1000</th>
<th>2500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\hat{\alpha}$)</td>
<td>1.0571</td>
<td>1.0203</td>
<td>1.0081</td>
</tr>
<tr>
<td>standard deviation($\hat{\alpha}$)</td>
<td>0.31</td>
<td>0.17</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 4. Performance of different estimates of the statistics $\hat{\alpha}$ of $\alpha = 0.5$ based on 5000 simulations.

<table>
<thead>
<tr>
<th>T-2</th>
<th>1000</th>
<th>2500</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\hat{\alpha}$)</td>
<td>0.512</td>
<td>0.503</td>
<td>0.501</td>
</tr>
<tr>
<td>standard deviation($\hat{\alpha}$)</td>
<td>0.09</td>
<td>0.06</td>
<td>0.04</td>
</tr>
</tbody>
</table>

• **The case** $\alpha > 1$. As we mentioned above, our estimator may not be good enough to estimate large values of $\alpha$. This is best illustrated in Table 5 where we report for the large variability of $\hat{\alpha}$ and the percentage of its negative values. For $\alpha = 10$, we report for the standard deviation and percentage of negative values of $\hat{\alpha}$ based on $N = 5000$ simulations drawn from the normal distribution of $Px$ for $T - 2 = 1000$, 7000 and 14000. The reported figures show large values of the standard deviation of $\hat{\alpha}$ and that $25\%$, $4.4\%$ respectively $1\%$ of its values are negative. Although these values decrease as the time horizon $T$ gets larger, we would like to be cautious when using it to estimate large values of $\alpha$.

For the sake of completeness, we include in Table 5 the performance of the associated statistics $\hat{\beta}$.
Table 5. Performance of different estimates of the statistics $\hat{\alpha}$ of $\alpha = 10$ based on 5000 simulations.

<table>
<thead>
<tr>
<th>T-2</th>
<th>1000</th>
<th>7000</th>
<th>14000</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentage of $(\hat{\alpha} &lt; 0)$</td>
<td>0.255</td>
<td>0.044</td>
<td>0.010</td>
</tr>
<tr>
<td>standard deviation($\hat{\alpha}$)</td>
<td>548</td>
<td>226</td>
<td>65</td>
</tr>
<tr>
<td>mean($\hat{\beta}$)</td>
<td>6.112</td>
<td>6.1018</td>
<td>6.1008</td>
</tr>
<tr>
<td>standard deviation($\hat{\beta}$)</td>
<td>0.167</td>
<td>0.063</td>
<td>0.044</td>
</tr>
</tbody>
</table>

4.2. Calibration of the trend from some market data

The market data used for this exercise include:
- The quarterly US real GNP under the period January 1, 1947 – January 1, 2006;
- The monthly British Pound spot price (GBP) under the period January 29, 1971 – July 31, 2006;
- The monthly Euro spot price (Euro) under the period December 31, 1998 – July 31, 2006;
- The monthly Swiss Franc spot price (CHF) under the period January 29, 1971 – July 31, 2006;
- The monthly Japanese Yen spot price (JPY) under the period January 29, 1971 – July 31, 2006;
- The monthly S&P 500 spot price (S&P) under the period August 31, 1956 – July 31, 2006;
- The monthly Nasdaq 100 spot price (Nasdaq) under the period February 28, 1985 – July 31, 2006;
- The monthly Dow Jones Euro stoxx 500 spot price (DJ) under the period December 31, 1985 – July 31, 2006;
- The monthly FTSEI 100 spot price (FTSI) under the period January 31, 1984 – July 31, 2006;

In Tables 6 and 7, we collect the estimated values of the statistics $\hat{\alpha}$, based on our estimator and $\alpha^*$ using the estimator suggested in Schlicht (2005). The values of $\alpha^*$ have been kindly communicated to us by E. Schlicht. We note that both approaches yielded estimated values of $\alpha$, roughly of comparable magnitude.

As shown for example in Figures 1 and 2 below, we get a quite good fit of the trend, of the considered US real GNP time serie, based on both Schlicht’s and our estimators, where the residual component is white noise, while the one corresponding to the Hodrick-Prescott filter ($\alpha = 1600$) is not.

Dislike Reeves et al. (2000) and Araujo et al. (2003), that found the third order shift operator more appropriate than the second order one to extract the trend from some Foreign Exchange Rate series, our estimator of the trend shows a good fit, using only the second order shift operator.

Table 6. Estimated values of the smoothing parameter $\alpha$. The values of $\hat{\alpha}$ are based on the estimator given by Eq. (3.11) and $\alpha^*$ are obtained using the Schlicht’s estimator.

<table>
<thead>
<tr>
<th>x</th>
<th>US real GNP</th>
<th>GBP</th>
<th>Euro</th>
<th>CHF</th>
<th>JPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.4137</td>
<td>0.3734</td>
<td>0.3675</td>
<td>0.5896</td>
<td>0.4489</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>2.4028</td>
<td>1.2998</td>
<td>0.6713</td>
<td>1.4592</td>
<td>1.8383</td>
</tr>
</tbody>
</table>
Table 7. Estimated values of the smoothing parameter $\alpha$. The values of $\hat{\alpha}$ are based on the estimator given by Eq. (3.11) and $\alpha^*$ are obtained using the Schlicht’s estimator.

<table>
<thead>
<tr>
<th>x</th>
<th>S&amp;P 500</th>
<th>Nasdaq</th>
<th>Dow Jones</th>
<th>FTSE</th>
<th>Nikkei</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.4963</td>
<td>0.5152</td>
<td>0.6417</td>
<td>0.4125</td>
<td>0.4721</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>8.1615</td>
<td>3.2756</td>
<td>1.1907</td>
<td>3.0841</td>
<td>2.4028</td>
</tr>
</tbody>
</table>

*************** Figure 1 about here ******************

*************** Figure 2 about here ******************

Appendix

Proof of Proposition 3. We first show that for $k = 0, 1, T \geq 4$, and $0 \leq \lambda \leq 2a_k/3\|A_kV\|_\infty$, we have

$$P(|\hat{r}_k - r_k| \geq a_k\lambda) \leq \exp(-\frac{\lambda^2}{4}).$$

Indeed, we use the following general result for Gaussian quadratic forms (see e.g. Giurcanu and Spokoiny (2002)). Let $Y$ be a centered Gaussian vector with covariance matrix $C$. Then it holds that

$$P(|Y^tAY - \mathbb{E}(Y^tAY)| \geq \lambda\sqrt{\text{var}(Y^tAY)}) \leq \max\{\exp(-\frac{\lambda^2}{4}), \exp(-\frac{\lambda\sqrt{\text{var}(Y^tAY)}}{6\|CA\|_\infty})\}.$$  

Therefore, from the representation $\hat{r}_k = (Px)^tA_kPx$ we get that

$$P(|(Px)^tA_kPx - \mathbb{E}((Px)^tA_kPx)| \geq \lambda a_k) \leq \max\{\exp(-\frac{\lambda^2}{4}), \exp(-\frac{\lambda a_k}{6\|VA_k\|_\infty})\}.$$  

Now we precisely estimate the quantity $\|VA_k\|_\infty$. Indeed, we have

$$A_0V = \frac{1}{(T-2)}V.$$  

Hence,

$$\|A_0V\|_\infty = \frac{r_0}{T-2}.$$  

Now,

$$A_1V(i, j) = \frac{1}{2(T-3)}[V(i - 1, j) + V(i + 1, j)]$$  

implies that

$$\|A_1V\|_\infty = (T - 3)\max(r_0, r_2, 2|r_1|).$$  

Using these estimate we get the results, noting that the third estimate follows from the fact that $\max(a_0^2, a_1^2) \leq \frac{10r^2}{r_0^2}$. 

Acknowledgements

Many thanks to Ekkehart Schlicht for insightful discussions and for his help to compute the values of his estimator of the smoothing parameter, reported in Table 3 above, for our dataset, using his Mathematica package for Seasonal Adjustment. We would like to thank the referees for many helpful and valuable comments and suggestions.

References

of Monetary Economics, 18: 49-75.


Figure 1. The estimated trend fitted with the quarterly US real GNP under the period 01-01-1947 to 01-01-2006. The Hodrick-Prescott trend corresponds to $\alpha = 1600$.

Figure 2. The autocorrelation function of the cyclical component (the noise term) of the quarterly US real GNP under the period between 01-01-1947 and 01-01-2006, using the value $\alpha = 1600$ suggested by the ones corresponding to the Schlicht’s and our estimators, reported in Table 3.