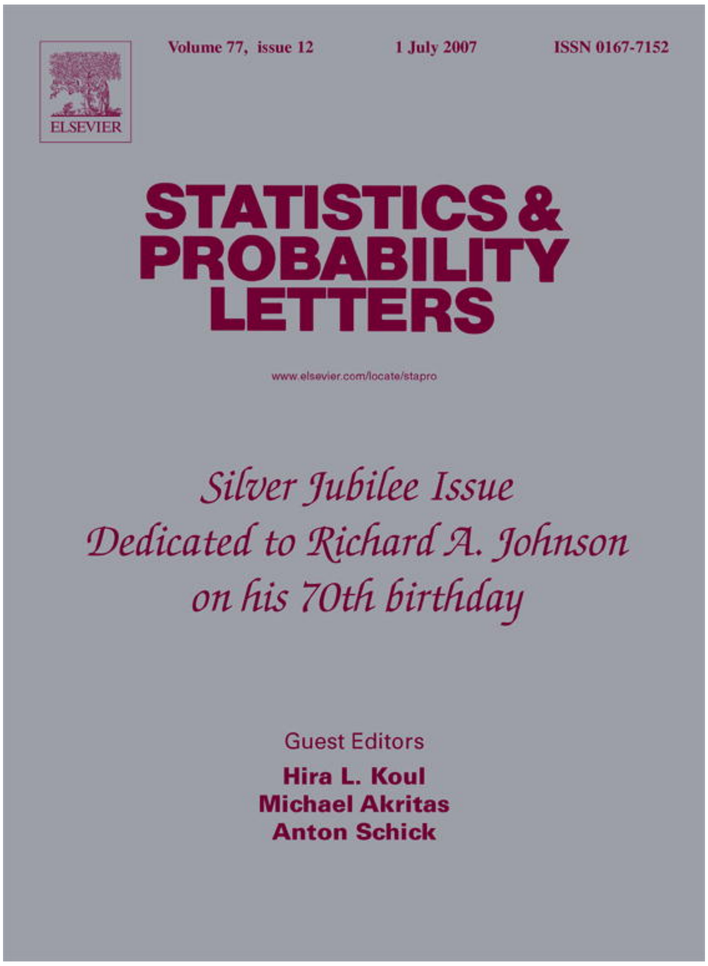

Volume 77, issue 12 1 July 2007 ISSN 0167-7152

ELSEVIER

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Silver Jubilee Issue
Dedicated to Richard A. Johnson
on his 70th birthday

Guest Editors
Hira L. Koul
Michael Akritas
Anton Schick

Available online at www.sciencedirect.com

ScienceDirect

Statistics & Probability Letters 77 (2007) 1225–1234

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Time series smoothing by penalized least squares

Victor M. Guerrero

Departamento de Estadística, Instituto Tecnológico Autónomo de México (ITAM), México 01000, D.F., Mexico

Available online 16 March 2007

Abstract

The time series smoothing problem is approached in a slightly more general form than usual. The proposed statistical solution involves an implicit adjustment to the observations at both extremes of the time series. The resulting estimated trend becomes more statistically grounded and an estimate of its sampling variability is provided. An index of smoothness is derived and proposed as a tool for choosing the smoothing constant.

Keywords: Mean square error; Smoothness index; Unobserved components

1. Introduction

A basic objective of a time series analysis is to estimate unobserved components that are assumed to underlie the time series under study. In fact, the following unobserved components model of an observable time series provides a useful and simple representation of the data $\{Z_t\}$:

$$Z_t = \tau_t + \eta_t \quad \text{for } t = 1, \ldots, N, \tag{1}$$

where τ_t denotes the trend or signal of the data and η_t is the noise at time t. Researchers in different fields have employed this representation. In an early use of (1), Whittaker (1923) and Henderson (1924) suggested to graduate (smooth) actuarial data by solving the following penalized least squares problem for $\mu = 0$ and $\lambda > 0$:

$$\min_{\{\tau_t\}} \sum_{t=1}^{N} (Z_t - \tau_t)^2 + \lambda \sum_{t=d+1}^{N} (\nabla^d \tau_t - \mu)^2. \tag{2}$$

Here, μ denotes the mean, if it exists, or just a reference level for $\{\nabla^d \tau_t\}$, where $\nabla^d \tau_t$ denotes the dth order difference of $\{\tau_t\}$, with d a nonnegative integer. That is, $\nabla^0 \tau_t = \tau_t$, $\nabla \tau_t = \tau_t - \tau_{t-1}$, $\nabla^2 \tau_t = \nabla(\nabla \tau_t)$ and so on, in such a way that the second term of (2) is related to the smoothness of $\{\tau_t\}$.

The constant λ is called the smoothing parameter since it penalizes the lack of smoothness in the trend. That is, $\{\tau_t\} \to \{Z_t\}$ as $\lambda \to 0$, in which case the smoothness diminishes, while as $\lambda \to \infty$, $\{\tau_t\}$ gets closer to the (smooth) polynomial implied by $\nabla^d \tau_t = \mu$. Thus, in the latter case, when $d = 0$ the trend will be constant and when $d \geqslant 1$ the trend will be given by $\tau_t = \beta_0 + \beta_1 t + \cdots + \beta_{d-1} t^{d-1} + (\mu/d!) t^d$, where the constants β_i, for

E-mail address: guerrero@itam.mx.

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2007.03.006

$i = 0, \ldots, d-1$, depend on the first d values of $\{\tau_t\}$. Therefore, assuming $\mu = 0$ has implications on the degree of the polynomial.

The value of d in (2) is usually chosen by the analyst on a priori grounds as $d \leqslant 2$, and seldom is $d \geqslant 3$ used in practice. When $\mu = 0$ and $d = 1$ or 2, the corresponding solutions to the minimization problem are well known in the financial and econometric literature. They are called exponential smoothing and Hodrick–Prescott filtering, respectively (see King and Rebelo, 1993). When the observed data are not equally spaced the problem has been considered in the general setting of smoothing splines (see Wahba, 1990). However, the observations of a time series are equally spaced and the problem has also received considerable attention in this context, as is evidenced in Kitagawa and Gersch (1996) and Kaiser and Maravall (2001).

In what follows, a statistical solution to the smoothing problem is proposed. Then, the focus is placed on the problem of selecting a λ value in such a way that it produces a percentage of smoothness for the trend specified beforehand. This result is specialized to the cases $d = 0, 1$ and 2, which are considered of outmost practical interest. Some numerical examples are shown to provide empirical evidence of the usefulness of the method proposed here.

2. A statistical solution

The minimization problem (2) is slightly more general than the usual problem considered in the literature because μ is not assumed to be zero beforehand. A solution to the problem for $\mu = 0$ can be found in Kitagawa and Gersch (1996, p. 34) who approached it from a least squares computational perspective, while King and Rebelo (1993) employed optimal linear filtering tools to obtain essentially the same solution. In this paper, the problem is posed as the estimation of a random vector in order to develop a statistical solution and to derive an index of smoothness. Thus, let us consider the following *tentative* statistical model for $\{\tau_t\}$, similar to that used by Hodrick and Prescott (1997) or Kitagawa and Gersch (1996), that is:

$$\nabla^d \tau_t = \mu + \varepsilon_t \quad \text{for } t = d+1, \ldots, N, \tag{3}$$

with $\{\varepsilon_t\}$ a sequence of serially uncorrelated and identically distributed random errors with mean zero and $Var(\varepsilon_t) = \sigma_\varepsilon^2$.

Now, let us define the following arrays: $\mathbf{Z} = (Z_1, \ldots, Z_N)'$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_N)'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)'$ are $N \times 1$ vectors; $\boldsymbol{\varepsilon} = (\varepsilon_{d+1}, \ldots, \varepsilon_N)'$ and $\mathbf{1}_{N-d} = (1, \ldots, 1)'$ are $(N-d) \times 1$ vectors and K_d is the matrix representation of the difference operator ∇^d, so that

$$K_d = \begin{pmatrix} & \mathbf{k}_d & & \mathbf{0}_{N-d-1} \\ 0 & & \mathbf{k}_d & \mathbf{0}_{N-d-2} \\ & & \cdots & \\ & \mathbf{0}_{N-d-1} & & \mathbf{k}_d \end{pmatrix} \tag{4}$$

is an $(N-d) \times N$ matrix, with $\mathbf{k}_d$ the $1 \times (d+1)$ vector given by

$$\mathbf{k}_d = \left((-1)^d \binom{d}{d}, (-1)^{d-1} \binom{d}{d-1}, \ldots, (-1) \binom{d}{1}, \binom{d}{0} \right) \tag{5}$$

and $\mathbf{0}_m$ the $m \times 1$ zero vector. Thus, the vector of data $\mathbf{Z}$ can be expressed as

$$\mathbf{Z} = \boldsymbol{\tau} + \boldsymbol{\eta}, \tag{6}$$

where the trend component is given, for μ known, by

$$K_d \boldsymbol{\tau} = \mu \mathbf{1}_{N-d} + \boldsymbol{\varepsilon}, \tag{7}$$

with $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ random vectors such that $E(\boldsymbol{\eta}) = \mathbf{0}_N$, $\operatorname{Var}(\boldsymbol{\eta}) = \sigma_\eta^2 V$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}_{N-d}$, $\operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 I_{N-d}$ and $E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') = 0$, with V a known positive definite matrix.

The following result combines two sources of information in order to obtain an optimal linear predictor, in mean square error (MSE) sense, of a random vector.

Ese lema es el estimador lineal de MSE mínimo (Wiener–Gauss–Markov) —equivalente a una sola actualización de Kalman con "prior" y una medición

ver filtro de kalman version estatica

el lema es el "paso de actualización" de mínimo error cuadrático lineal (LMMSE)
se parte de un pronóstico/prior de X I W y lo corriges con lo que dice la nueva me
Y, ponderando por una ganancia óptima que equilibra "señal vs. ruido".

Lemma. *Let us suppose that an unobservable random vector* $\mathbf{X}$ *is related to the observable vectors* $\mathbf{W}$ *and* $\mathbf{Y}$ *by means of*

X el que queremos estimar
/gamma es w, incertidumbre
W unas condiciones ya conocidas como un prior

Y es una observacion con ruido de medicion

$$\mathbf{X} - E(\mathbf{X}|\mathbf{W}) = \gamma \quad \text{and} \quad \mathbf{Y} = C\mathbf{X} + \boldsymbol{\delta}, \tag{8}$$

where γ *is a random vector corresponding to a stationary process, with* $E(\gamma|\mathbf{W}) = \mathbf{0}$ *and* $E(\gamma\gamma'|\mathbf{W}) = P$. C *is a known full rank matrix and* $\boldsymbol{\delta}$ *is another random vector coming from a stationary process with* $E(\boldsymbol{\delta}|\mathbf{W}) = \mathbf{0}$, $E(\boldsymbol{\delta}\boldsymbol{\delta}'|\mathbf{W}) = R$ *and* $E(\gamma\boldsymbol{\delta}'|\mathbf{W}) = \mathbf{0}$. *Then, the minimum MSE linear estimator* (*predictor*) *of* $\mathbf{X}$ *based on* $\mathbf{W}$ *and* $\mathbf{Y}$ *becomes*

$$\hat{\mathbf{X}} = E(\mathbf{X}|\mathbf{W}) + A[\mathbf{Y} - CE(\mathbf{X}|\mathbf{W})]. \tag{9}$$

Its corresponding MSE matrix $\Sigma = \mathrm{Var}(\hat{\mathbf{X}} - \mathbf{X}|\mathbf{W})$ *is given by*

$$\Sigma = P(I - C'A'), \tag{10}$$

where

$$A = PC'(CPC' + R)^{-1}. \tag{11}$$

Proof. The Law of Iterated Projections (see Sargent, 1979, p. 208) leads us to

$$E(\mathbf{X}|\mathbf{W},\mathbf{Y}) = E(\mathbf{X}|\mathbf{W}) + E[\mathbf{X} - E(\mathbf{X}|\mathbf{W})|\mathbf{Y} - E(\mathbf{Y}|\mathbf{W})].$$

Then, from (8) we get $\mathbf{Y} - E(\mathbf{Y}|\mathbf{W}) = (C\mathbf{X} + \boldsymbol{\delta}) - CE(\mathbf{X}|\mathbf{W}) = C\gamma + \boldsymbol{\delta}$ and $E[\mathbf{X} - E(\mathbf{X}|\mathbf{W})|\mathbf{Y} - E(\mathbf{Y}|\mathbf{W})] = E(\gamma|C\gamma + \boldsymbol{\delta})$. Since we are looking for a linear estimator, the last expectation must be of the form $E(\gamma|C\gamma + \boldsymbol{\delta}) = A(C\gamma + \boldsymbol{\delta})$, with A a constant matrix. Moreover, to minimize the length of $\gamma - A(C\gamma + \boldsymbol{\delta})$, the following orthogonality condition must hold $E\{(C\gamma + \boldsymbol{\delta})[\gamma' - (\gamma'C' + \boldsymbol{\delta})A']|\mathbf{W}\} = 0$ and this yields (11). Therefore, given $\mathbf{W}$ and $\mathbf{Y}$ we get (9), while (10) follows from

$$\hat{\mathbf{X}} - \mathbf{X} = [E(\mathbf{X}|\mathbf{W}) + A(C\gamma + \boldsymbol{\delta})] - [E(\mathbf{X}|\mathbf{W}) + \boldsymbol{\delta}] = (AC - I)\gamma + A\boldsymbol{\delta}.$$

Hence, given $\mathbf{W}$ and $\mathbf{Y}$,

$$\begin{aligned}\mathrm{Var}(\hat{\mathbf{X}} - \mathbf{X}|\mathbf{W},\mathbf{Y}) &= (AC - I)P(C'A' - I) + ARA' \\ &= P - ACP - PC'A' + ACPC'A' + ARA' \\ &= P(I - C'A'),\end{aligned}$$

where the last equality holds because $ACP = PC'A' = A(CPC' + R)A'$. □

Since $E(\boldsymbol{\tau}|\mathbf{Z}) = \mathbf{Z}$, an application of the lemma with $\mathbf{X} = \boldsymbol{\tau}$, $\mathbf{W} = \mathbf{Z}$, $\gamma = -\boldsymbol{\eta}$, $\mathbf{Y} = \mu\mathbf{1}_{N-d}$, $C = K_d$, $\boldsymbol{\delta} = \boldsymbol{\varepsilon}$, $P = \sigma_\eta^2 V$ and $R = \sigma_\varepsilon^2 I_{N-d}$ produces the following result.

Proposition 1. *The minimum MSE linear estimator* (*predictor*) *of the trend* $\boldsymbol{\tau}$ *is*

$$\hat{\boldsymbol{\tau}} = (V^{-1} + \lambda K_d'K_d)^{-1}(V^{-1}\mathbf{Z} + \lambda\mu K_d'\mathbf{1}_{N-d}), \tag{12}$$

with MSE matrix

$$\Sigma = \sigma_\eta^2(V^{-1} + \lambda K_d'K_d)^{-1}. \tag{13}$$

Proof. From (9) and (10) we get $\hat{\boldsymbol{\tau}} = \mathbf{Z} + A(\mu\mathbf{1}_{N-d} - K_d\mathbf{Z})$, whose MSE matrix is $\Sigma = \sigma_\eta^2(I_N - AK_d)V$, where $A = \sigma_\eta^2 VK_d'(\sigma_\eta^2 K_d VK_d' + \sigma_\varepsilon^2 I_{N-d})^{-1}$. These results can be expressed in a more familiar form within the time series smoothing framework,

$$\begin{aligned}\Sigma &= \sigma_\eta^2 V - \sigma_\eta^2 VK_d'(\sigma_\eta^2 K_d VK_d' + \sigma_\varepsilon^2 I_{N-d})^{-1}K_d V\sigma_\eta^2 \\ &= (\sigma_\eta^{-2}V^{-1} + \sigma_\varepsilon^{-2}K_d'K_d)^{-1}\end{aligned}$$

and

$$\begin{aligned}\hat{\boldsymbol{\tau}} &= (\sigma_\eta^{-2}\Sigma V^{-1}\mathbf{Z} + \sigma_\varepsilon^{-2}\mu\Sigma K_d'\mathbf{1}_{N-d})^{-1} \\ &= (\sigma_\eta^{-2}V^{-1} + \sigma_\varepsilon^{-2}K_d'K_d)^{-1}(\sigma_\eta^{-2}V^{-1}\mathbf{Z} + \sigma_\eta^{-2}\mu K_d'\mathbf{1}_{N-d}).\end{aligned}$$

Furthermore, since the smoothing parameter is defined as $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$, the last two expressions become (13) and (12). $\square$

Remarks. (i) The lemma allows us to estimate a random vector coming from a nonstationary time series process. In fact, the first equation of (8) is valid both for stationary and nonstationary processes, as it was shown by Bell (1984). The second equation of (8) is valid only if the processes corresponding to $\mathbf{Y}$ and $C\mathbf{X}$ are either both stationary or else they are nonstationary but cointegrated. This fact is important from a statistical point of view in order to apply Proposition 1 appropriately, because $\mathbf{Y} = \mu\mathbf{1}_{N-d}$ being a constant vector implies that $C\mathbf{X} = K_d\boldsymbol{\tau}$ must correspond to a stationary process $\{\nabla^d\tau_t\}$. Hence, d must be chosen accordingly.

(ii) In a less formal way, the estimator $\hat{\boldsymbol{\tau}}$ can also be obtained from

$$\begin{pmatrix} \mathbf{Z} \\ \mu\mathbf{1}_{N-d} \end{pmatrix} = \begin{pmatrix} I_N \\ K_d \end{pmatrix}\boldsymbol{\tau} + \begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \end{pmatrix}, \tag{14}$$

with

$$E\begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \end{pmatrix} = \mathbf{0}_{2N-d} \quad \text{and} \quad \operatorname{Var}\begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma_\eta^2 V & 0 \\ 0 & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}, \tag{15}$$

in which case generalized least squares (GLS) produces (12) and (13). Thus, the intuitive interpretation of GLS carries over to the results provided by Proposition 1.

(iii) Expression (12) is clearly a generalization of existing work. If we let $\mu = 0$, then appropriate specification of the matrix V allows the proposed model to take into account heteroskedasticity and autocorrelation of the observed series. For instance, Reeves et al. (2000) consider a diagonal matrix V with nonconstant variances σ_t^2, for $t = 1, \ldots, N$. Then, the resulting estimated trend corresponds to a time-varying smoothing parameter λ_t. The matrix V could also be used to account for the presence of cycles in the observed series by considering an appropriate autocorrelation structure (e.g. a moving average model of finite order or an auto-regressive model of order $p \geqslant 2$). However, in practice it is customary to make $V = I_N$ and that is the case considered below. Even in such a case, by letting d be a nonnegative integer we have as special cases of (12) the well-known exponential smoothing ($d = 1$) and Hodrick–Prescott ($d = 2$) filters, studied by King and Rebelo (1993).

(iv) When $d \geqslant 1$, the vector $K_d'\mathbf{1}_{N-d}$ is zero except for its first d and last d elements. This follows because $\sum_{i=0}^{d}(-1)^i\binom{d}{i} = 0$, so that

$$K_d'\mathbf{1}_{N-d} = \left(\sum_{i=d}^{d}(-1)^i\binom{d}{i}, \ldots, \sum_{i=1}^{d}(-1)^i\binom{d}{i}, 0, \ldots, 0, \sum_{i=0}^{d-1}(-1)^i\binom{d}{i}, \ldots, \sum_{i=0}^{0}(-1)^i\binom{d}{i}\right)'. \tag{16}$$

Therefore, the observed values of $\{Z_t\}$ enter the formula of the estimator $\hat{\boldsymbol{\tau}}$ modified in both of its extremes by the mean value μ weighted by λ. On the other hand, if $d = 0$ the estimator becomes $\hat{\boldsymbol{\tau}} = \alpha\mathbf{Z} + (1-\alpha)\mu\mathbf{1}_N$, with $\alpha = (1+\lambda)^{-1}$.

(v) The effect of the mean should be kept in mind when extrapolating the trend, since $\mu \neq 0$ implies the trend follows a polynomial of degree d, while $\mu = 0$ implies a polynomial of degree $d-1$. Furthermore, the extrapolated values will depend critically on the last d estimated trend values. That is, if we call $\hat{\tau}_N(h)$ the h-period ahead forecast of τ_{N+h}, with origin at N, then for $h \geqslant 1$we get $\hat{\tau}_N(h) = \mu$ if $d = 0$, $\hat{\tau}_N(h) = h\mu + \tau_N$ if $d = 1$ and $\hat{\tau}_N(h) = [h(h+1)/2]\mu + (h+1)\tau_N - h\tau_{N-1}$ if $d = 2$.

A feasible trend estimator must take into account the fact that μ is commonly unknown and therefore it has to be estimated from the very data. Thus, an unbiased estimator of μ is given by the sample mean of $\{\nabla^d Z_t\}$, that is,

$$\hat{\mu} = (N-d)^{-1}\mathbf{1}_{N-d}'K_d\mathbf{Z}. \tag{17}$$

So, expression (12) becomes

$$\hat{\boldsymbol{\tau}} = (I_N + \lambda K_d'K_d)^{-1}[I_N + \lambda(N-d)^{-1}K_d'\mathbf{1}_{N-d}\mathbf{1}_{N-d}'K_d]\mathbf{Z}. \tag{18}$$

Further, in order to measure variability around the estimated series $\{\hat{\tau}_t\}$ we need to estimate Σ in (13), which in turn requires an estimator of σ_η^2. This estimator is provided by the following result.

Proposition 2. *Let the assumptions of Proposition* 1 *be valid. Then, an unbiased estimator of* σ_η^2 *is given by*

$$\tilde{\sigma}_\eta^2 = [\mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\tau}}'(I_N + \lambda K_d' K_d)\hat{\boldsymbol{\tau}}]/(N-d) + \lambda\mu^2. \tag{19}$$

Proof. The result follows by writing

$$\mathrm{Var}\begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \end{pmatrix} = \sigma_\eta^2 \begin{pmatrix} I_N & 0 \\ 0 & \lambda^{-1} I_{N-d} \end{pmatrix} = \sigma_\eta^2 \Omega$$

and defining the residual vector as

$$\mathbf{e}^* = \begin{pmatrix} \hat{\boldsymbol{\eta}} \\ -\hat{\boldsymbol{\varepsilon}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} - \begin{pmatrix} I_N \\ K_d \end{pmatrix} \hat{\boldsymbol{\tau}}$$

with $\hat{\boldsymbol{\eta}} = \mathbf{Z} - \hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\varepsilon}} = K_d \hat{\boldsymbol{\tau}} - \mu \mathbf{1}_{N-d}$. Then, we get

$$\mathbf{e}^* = (I_{2N-d} - M)\begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \end{pmatrix},$$

where the matrix M is given by

$$M = \begin{pmatrix} I_N \\ K_d \end{pmatrix} (I_N + \lambda K_d' K_d)^{-1} (I_N \lambda K_d').$$

This is an idempotent matrix with trace $\mathrm{tr}(M) = N$. Therefore, the sum of squared residuals is given by

$$\mathbf{e}^{*\prime} \Omega^{-1} \mathbf{e}^* = \hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} + \lambda \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\tau}}'(I_N + \lambda K_d' K_d)\hat{\boldsymbol{\tau}} + (N-d)\lambda\mu^2.$$

Hence, an unbiased estimator of σ_η^2 becomes (19). □

Remark. By wrongly assuming $\mu = 0$, not only the endpoints of the estimator $\hat{\boldsymbol{\tau}}$ will be affected, but also the variance σ_η^2 will be underestimated by an amount that grows as $\lambda\mu^2$.

It should be stressed that the estimator $\tilde{\sigma}^2$ was obtained on the assumption that μ was a known parameter, while in fact it has to be estimated. Thus, when using $\hat{\mu}$ in place of μ, it makes sense to correct (19) for this fact. In that case, the proposed estimator is given by $\hat{\sigma}_\eta^2 = (N-d)\tilde{\sigma}_\eta^2/(N-d-1)$, that is,

$$\hat{\sigma}_\eta^2 = \left[\sum_{t=1}^{N} (Z_t - \hat{\tau}_t)^2 + \lambda \sum_{t=d+1}^{N} (\nabla^d \hat{\tau}_t - \hat{\mu})^2 \right] \Bigg/ (N-d-1). \tag{20}$$

3. A measure of smoothness

The precision matrix of $\hat{\boldsymbol{\tau}}$ is defined as

$$\Sigma^{-1} = \sigma_\eta^{-2} I_N + \sigma_\varepsilon^{-2} K_d' K_d. \tag{21}$$

This matrix is composed by two precision matrices, $\sigma_\eta^{-2} I_N$ associated with model (6) for the observations and $\sigma_\varepsilon^{-2} K_d' K_d$ associated with model (7) for the smooth component of the series. Then, as in Guerrero et al. (2001) we can measure the precision contributed by the smooth component to the total precision. Such a measure was originally derived by Theil (1963) to quantify the proportion of a matrix P in $(P+Q)^{-1}$, where P and Q are $N \times N$ positive definite matrices. Theil's measure is

$$\Lambda(P; P+Q) = \mathrm{tr}[P(P+Q)^{-1}]/N. \tag{22}$$

This measure of relative precision has the following properties: (i) it lies in the interval $[0, 1]$; (ii) it is invariant under linear non-singular transformations of the variable involved; (iii) it behaves linearly and (iv) $\Lambda(P; P+Q) + \Lambda(Q; P+Q) = 1$.

Table 1
Values of λ for selected values of d and $S_d(\lambda, N)$, with $N = 100$

Difference	$S_d(\lambda, N)$				
d	0.5	0.6	0.7	0.8	0.9
0	1.000	1.500	2.333	4.000	9.000
1	0.765	1.346	2.614	6.312	27.420
2	0.427	0.970	2.812	13.506	244.872

The proposal is to use (22) to measure the proportion of precision induced by the smoothness component. The corresponding smoothness index is given by

$$S_d(\lambda, N) = \Lambda(\sigma_\varepsilon^{-2} K_d' K_d; \Sigma^{-1}) = \begin{cases} \lambda(1+\lambda)^{-1} & \text{if } d = 0, \\ 1 - \operatorname{tr}[(I_N + \lambda K_d' K_d)^{-1}]/N & \text{if } d \geqslant 1. \end{cases} \tag{23}$$

It is clear that $S_d(\lambda, N) \to 0$ as $\lambda \to 0$ and $S_d(\lambda, N) \to 1$ as $\lambda \to \infty$. Then, we should specify the smoothness $S_d(\lambda, N)$ and find the corresponding λ value for fixed values of N and d. If $d = 0$ we get $\lambda = S_0(\lambda, N)/[1 - S_0(\lambda, N)]$ and if $d \geqslant 1$ we can use a nonlinear routine to solve (23) for λ, given N and d. For instance, when $N = 100$, the λ values corresponding to different smoothness indices are shown in Table 1.

4. Some illustrative examples

In this section, two examples are provided to shed some light into the proposed procedure and the results that can be obtained through its application in practice.

Example 1. Monthly mean temperature (°C) in December for a region of the State of Veracruz, Mexico. The region of reference is geographically located at −98 to −93 longitude and 17–22 latitude. A study of this kind of data is generally done to assess the potential impacts of climate change on human activities and natural systems. It is a common belief that the climate is changing and, therefore, it is reasonable to estimate trends of weather series (see, for instance, Jewson and Penzer, 2006). The data employed in this illustration cover the years 1901–1995 as shown in Table 2 and their study is important to assess the effect of climate change on coffee production.

A graph of the temperature series shows an underlying slow changing mean pattern (see Fig. 1) that indicates that the series may be stationary. An application of the augmented Dickey–Fuller (ADF) unit root test confirms the idea that $d = 0$. The estimated ADF regression employed was (standard errors in parenthesis)

$$\begin{aligned} \nabla Z_t = {} & \underset{(3.14)}{13.42} - \underset{(0.14)}{0.62} Z_{t-1} - \underset{(0.13)}{0.09} \nabla Z_{t-1} - \underset{(0.10)}{0.02} \nabla Z_{t-2} \end{aligned}$$

so that the statistic $\tau_\mu = -4.26$ leads to rejecting the null hypothesis of a unit root by comparing it with the asymptotic 1% critical point given by −3.43.

The results of applying the smoothing procedure are shown in Figs. 1 and 2. We see that the smoother the trend, the smaller the uncertainty around it. Thus, as the degree of smoothness increases, less data points are included within the two-standard error limits.

Example 2. Smoothing and trend forecasting of Mexico's real gross domestic product (GDP). Quarterly data on seasonally adjusted GDP from 1980:1 to 2006:2 were obtained from the website of the National Institute of Statistics, Geography and Informatics (INEGI, www.inegi.gob.mx). Table 3 shows the data for all quarters (Q1, …, Q4) from 1980:1 up to 2005:4 employed for smoothing, the remaining two data points will be used to check the validity of the forecasts.

Table 2
Mean temperature of December in a region of the State of Veracruz, Mexico

Years	Temperature in °C									
1901–1910	21.68	21.12	19.96	20.00	19.66	20.52	20.98	21.54	21.72	20.16
1911–1920	21.76	21.60	21.38	22.32	22.40	22.68	20.76	21.38	21.70	22.48
1921–1930	21.94	21.80	22.04	21.78	20.72	22.54	22.24	21.44	21.18	20.96
1931–1940	22.66	22.26	21.92	22.16	21.54	21.44	21.62	20.64	22.14	22.24
1941–1950	23.10	22.48	21.04	20.50	21.84	21.84	20.94	22.52	21.82	20.44
1951–1960	22.86	22.54	22.90	22.10	22.80	23.04	22.26	22.54	22.26	20.86
1961–1970	22.36	21.72	20.24	21.92	21.22	20.62	22.30	21.48	21.74	22.48
1971–1980	23.48	21.92	20.52	22.02	20.92	20.36	22.14	22.40	21.40	20.48
1981–1990	22.14	22.02	21.92	22.66	21.86	21.58	22.32	21.82	19.60	21.58
1991–1995	22.08	23.20	22.08	22.70	22.26					

Source: IPCC Data Distribution Center, http://ipcc-ddc.cru.uea.ac.uk/java/time_series.html.

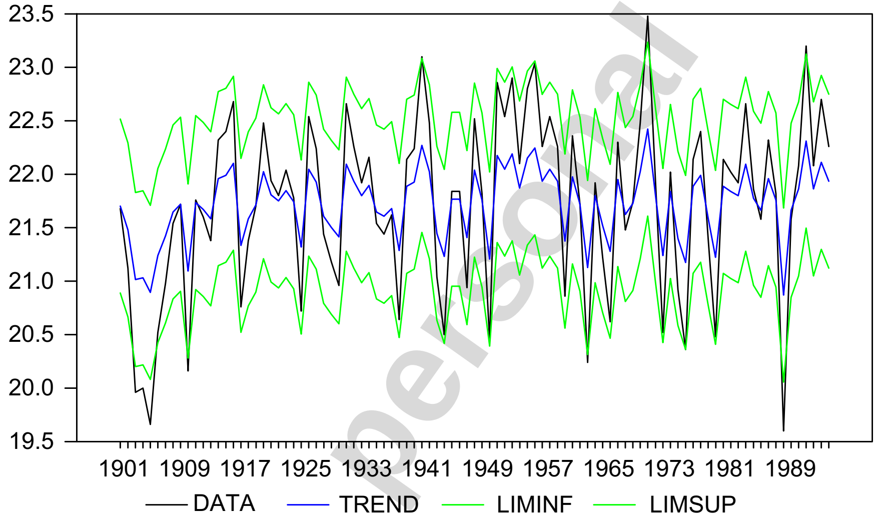

Fig. 1. Observed temperature, trend with 60% smoothness and two-standard error limits.

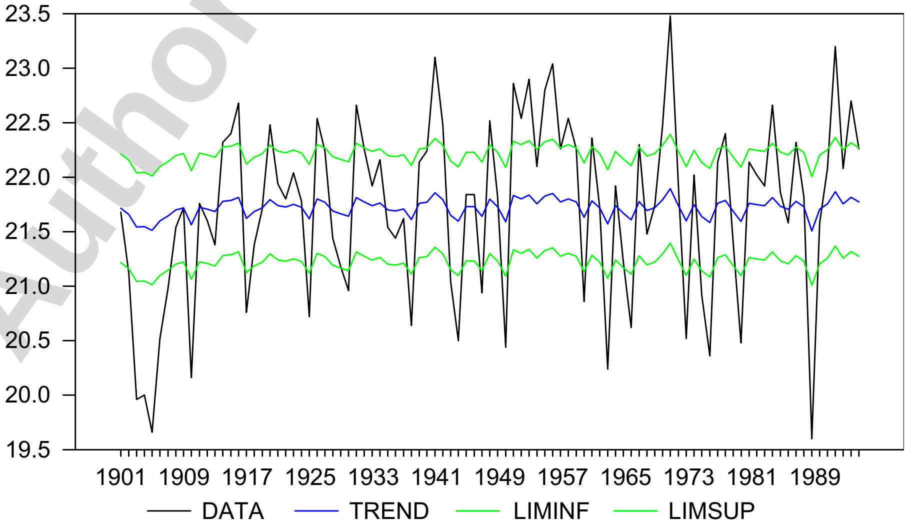

Fig. 2. Observed temperature, trend with 90% smoothness and two-standard error limits.

Table 3
Mexico's seasonally adjusted real GDP (millions of pesos at constant prices of 1993)

Year	Q1	Q2	Q3	Q4	Year	Q1	Q2	Q3	Q4
1980	927,175	933,282	953,408	981,005	1993	1,248,397	1,243,247	1,263,081	1,269,349
1981	1,002,873	1,028,945	1,034,287	1,052,803	1994	1,285,127	1,307,914	1,320,085	1,334,517
1982	1,032,472	1,034,157	1,026,666	1,004,127	1995	1,271,536	1,197,052	1,212,502	1,240,097
1983	996,799	976,591	984,778	996,096	1996	1,272,751	1,276,343	1,296,591	1,328,347
1984	1,023,495	1,008,657	1,032,948	1,023,840	1997	1,348,818	1,367,810	1,389,749	1,417,996
1985	1,042,443	1,041,846	1,048,176	1,045,330	1998	1,434,459	1,445,439	1,458,317	1,458,454
1986	1,025,365	1,021,222	1,001,958	999,218	1999	1,472,412	1,491,061	1,518,099	1,538,944
1987	1,005,031	1,032,919	1,033,814	1,046,291	2000	1,579,909	1,604,483	1,621,328	1,613,068
1988	1,039,207	1,035,951	1,037,192	1,057,847	2001	1,613,188	1,605,815	1,598,076	1,591,576
1989	1,077,778	1,077,952	1,098,202	1,088,509	2002	1,597,918	1,615,749	1,624,288	1,622,711
1990	1,111,879	1,136,521	1,151,835	1,166,461	2003	1,616,953	1,634,614	1,640,865	1,655,974
1991	1,169,716	1,187,390	1,189,810	1,211,274	2004	1,676,417	1,696,390	1,713,939	1,735,402
1992	1,210,803	1,231,241	1,242,485	1,243,492	2005	1,738,030	1,732,358	1,771,762	1,781,799

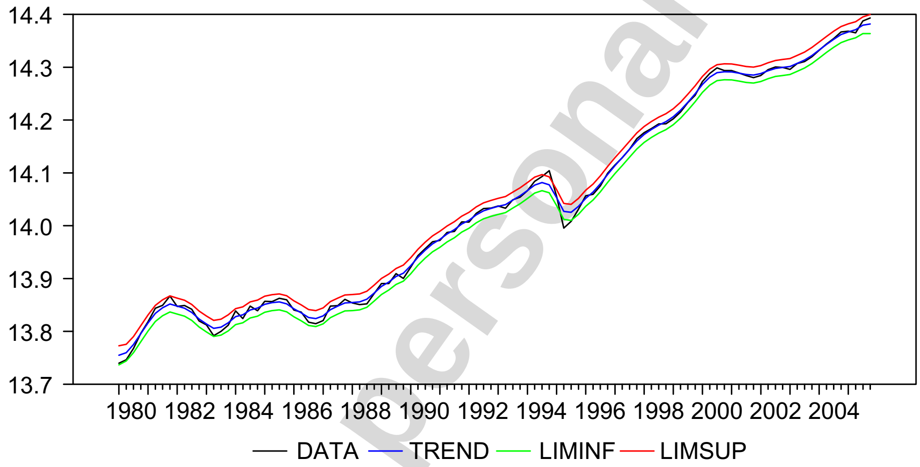

Fig. 3. Mexico's GDP (in logs), trend with 60% smoothness and two-standard error limits.

In order to perform business cycle analysis, GDP data are usually logged before applying the Hodrick–Prescott filter, which amounts to choosing $d = 2$ on a priori grounds. The logarithmic transformation was also applied here, but a unit root test was used to select the value of d empirically. The ADF regression model for $Z_t = \log(\mathrm{GDP}_t)$ became in this case

$$\begin{aligned}\nabla^2 Z_t = {} & 0.005 - 0.798\nabla Z_{t-1} + 0.066\nabla^2 Z_{t-1} + 0.236\nabla^2 Z_{t-2} \\ & (0.002) \quad (0.132) \qquad\quad (0.122) \qquad\quad (0.099).\end{aligned}$$

Since the ADF test statistic took on the value $\tau_\mu = -6.05$ we were led to reject the unit root hypothesis at the 1% significant level (the corresponding critical value is -3.43). Therefore, the value $d = 1$ was used in the smoothing procedure. The smoothing constant corresponding to $N = 104$ for a trend with 60% percentage of smoothness became $\lambda = 1.31$ and the estimated value $\hat{\sigma}_\eta = 0.0119$ was used to calculate two-standard error limits around the estimated trend shown in Fig. 3.

We can get forecasts of the trend by using the estimated mean of the series $\{\nabla \log(\mathrm{GDP}_t)\}$ and the last estimated trend value, that is, $\hat{\mu} = 0.0063$ and $\hat{\tau}_{2005:4} = 14.3818$. Thus, trend forecasts for quarters 2006:1 and 2006:2 became $\hat{\tau}_{2005:4}(1) = \hat{\mu} + \hat{\tau}_{2005:4} = 14.3881$ and $\hat{\tau}_{2005:4}(2) = 2\hat{\mu} + \hat{\tau}_{2005:4} = 14.3944$. On the other hand, if we apply the smoothing procedure to the whole sample (from 1980:1 to 2006:2) we get the following estimated trend figures with two-standard error intervals, for quarters 2005:4, 2006:1 and 2006:2, 14.3937 (14.3786, 14.4089), 14.4034 (14.3878, 14.4190) and 14.4086 (14.3906, 14.4266). In all three cases, these intervals cover the

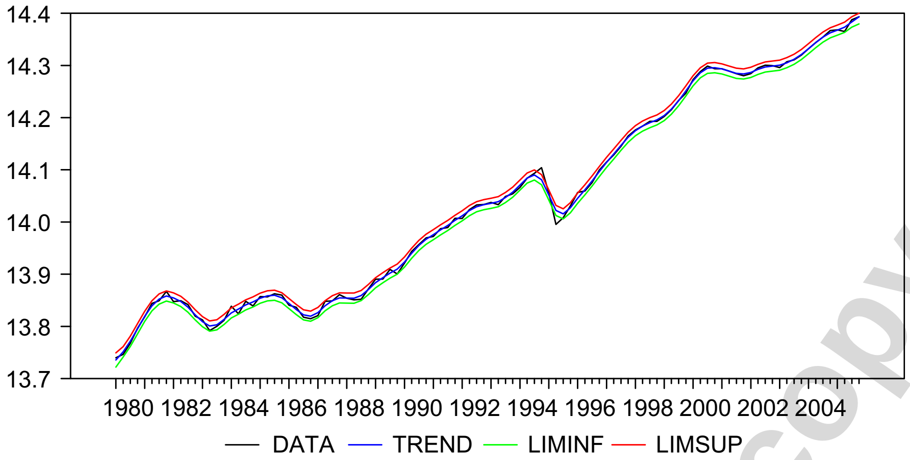

Fig. 4. Mexico's GDP (in logs), trend with $d = 2$, 60% smoothness and two-standard error limits.

previously estimated trend values. The trend estimates and forecasts in the original scale can be obtained by exponentiating the forecasts of the series in logs. In case we want to interpret the resulting forecasts as expected values, an adjustment should be applied to correct for bias introduced by this retransformation, as indicated in Guerrero (1993). Otherwise, if no adjustment is applied the forecasts are still valid, but they should be interpreted as estimated median values.

If we use $d = 2$, the smoothing constant for 60% of smoothness becomes $\lambda = 0.96$, and we get $\hat{\sigma}_\eta = 0.0077$. The corresponding graph is shown in Fig. 4. Trend forecasts can be obtained by using the estimated mean of the series $\{\nabla^2 \log(\mathrm{GDP}_t)\}$, $\hat{\mu} = -9 \times 10^{-6}$, as well as the two most recent estimated trend values, $\hat{\tau}_{2005:3} = 14.3832$ and $\hat{\tau}_{2005:4} = 14.3931$. Trend forecasts for 2006:1 and 2006:2 are $\hat{\tau}_{2005:4}(1) = \hat{\mu} + 2\hat{\tau}_{2005:4} - \hat{\tau}_{2005:3} = 14.4030$ and $\hat{\tau}_{2005:4}(2) = 3\hat{\mu} + 3\hat{\tau}_{2005:4} - 2\hat{\tau}_{2005:3} = 14.4129$.

In this illustration it is clear that the choice of d is crucial for trend estimation and forecasting. When $d = 1$, the results are conservative both in terms of variability and estimated trend values (particularly at the extremes of the series). On the contrary, if $d = 2$, the estimated trend is more volatile and the forecasts of future trend values become more adaptive. Thus, the degree of smoothness is not comparable for different values of d. Therefore we should select d in an objective (preferably data-based) way. Even if the values of d and λ have been established by the standard application of the smoothing method (e.g. the Hodrick–Prescott filter for quarterly series uses $d = 2$ and $\lambda = 1600$), we should be aware that the smoothness achieved varies according to the sample size. For instance, the smoothness achieved by applying the filter to the GDP data with $N = 104, d = 2$ and $\lambda = 1600$ is about 93%, whereas $\lambda = 1600$ produces only 88% of smoothness when $N = 20$ and $d = 2$.

5. Concluding remarks

The basic proposal of this work is to select the percentage of smoothness at the outset, instead of the smoothing constant. It is argued that the results obtained with the same degree of smoothness will be more comparable for different series or for the same series with different lengths. If we approach the time series smoothing problem from the standpoint suggested here, including the mean of the series in differences, we can get results that are not only justified from a computational perspective, but they are also well grounded in statistical theory. The basic theoretical results here derived provide an elementary justification of the proposed procedure, but more inferential tools are still required, perhaps based on classical normal distribution theory (e.g. to test whether $\mu = 0$ is a valid assumption).

Acknowledgements

The author is grateful to Asociación Mexicana de Cultura, A.C. for providing financial support to carry out this work through a Professorship on Time Series Analysis and Forecasting in Econometrics. He also thanks José L. Farah and an anonymous referee for providing useful comments on a previous version of this paper.

References

Bell, W.R., 1984. Signal extraction for nonstationary time series. Ann. Statist. 12, 646–664.

Guerrero, V.M., 1993. Time series analysis supported by power transformations. J. Forecasting 12, 37–48.

Guerrero, V.M., Juarez, R., Poncela, P., 2001. Data graduation based on statistical time series methods. Statist. Probab. Lett. 52, 169–175.

Henderson, R., 1924. A new method of graduation. Trans. Actuarial Soc. Amer. 25, 29–40.

Hodrick, R.J., Prescott, E.C., 1997. Postwar U.S. business cycles: an empirical investigation. J. Money Credit and Banking 29, 1–16.

Jewson, S., Penzer, J., 2006. Estimating trends in weather series: consequences for pricing derivatives. Stud. Nonlinear Dynamics and Econom. 10 (3) Article 9.

Kaiser, R., Maravall, A., 2001. Measuring Business Cycles in Economic Time Series. Lecture Notes in Statistics, vol. 154, Springer, New York.

King, R.G., Rebelo, S.T., 1993. Low frequency filtering and real business cycles. J. Econom. Dynamics Control 17, 207–231.

Kitagawa, G., Gersch, W., 1996. Smoothness Priors Analysis of Time Series. Lecture Notes in Statistics, vol. 116. Springer, New York.

Reeves, J.J., Blyth, C.A., Triggs, C.M., Small, J.P., 2000. The Hodrick–Prescott filter a generalization and a new procedure for extracting an empirical cycle from a series. Stud. Nonlinear Dynamics and Econom. 4 (1) Article 1.

Sargent, T.J., 1979. Macroeconomic Theory. Academic Press, New York.

Theil, H., 1963. On the use of incomplete prior information in regression analysis. J. Amer. Statist. Assoc. 58, 401–414.

Wahba, G., 1990. Spline Models for Observational Data. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Whittaker, E.T., 1923. On a new method of graduation. Proc. Edinburgh Math. Soc. 41, 63–75.