J. Japan Statist. Soc.
Vol. 35 No. 1 2005 99–119

ESTIMATING THE SMOOTHING PARAMETER IN THE SO-CALLED HODRICK-PRESCOTT FILTER

Ekkehart Schlicht*

This note gives a statistical description of the Hodrick-Prescott Filter (1997), originally proposed by Leser (1961). A maximum-likelihood estimator is derived and a related moments estimator is proposed that has a straightforward intuitive interpretation and coincides with the maximum-likelihood estimator for long time series. The method is illustrated by an application and several simulations. The statistical treatment in the state-space tradition implies some scepticism regarding the interpretation in terms of low-frequency filtering.

Key words and phrases: Adaptive estimation, Hodrick-Prescott filter, Kalman-Bucy, Kalman filtering, orthogonal parametrization, random walk, seasonal adjustment, spline, state-space models, time-series, time-varying coefficients, trend, Whittaker-Henderson graduation.

1. Introduction

What is known as the Hodrick-Prescott Filter (1997) is widely used in applications and has been embodied in various statistical packages. King and Rebelo (1993) write that the filter "is commonly used in investigations of the stochastic properties of real business cycle models," and many papers have been published that either use or improve the filter. The economics data bank EconLit lists seventy-two papers with "Hodrick-Prescott Filter" in title or abstract, and the statistics program packages eViews and Stata provide the filter as a standard feature.

The filter has been proposed originally by Leser (1961), building on the graduation method developed by Whittaker (1923) and Henderson (1924). It requires a smoothing constant as an input. This constant is usually fixed in an *ad hoc* way. The program eViews recommends 100 for annual data, 1600 for quarterly data, and 14.400 for monthly data, for instance, presumably summing up various findings in simulation studies and applied research. A theoretical approach to the determination of the filter has been suggested by Hodrick and Prescott (1997) who referred to Kalman-filtering and related the smoothing constant to a ratio of variances. (Their guess of a variance ratio of 1600 for quarterly data established a custom.) Earlier, Akaike (1980) and Schlicht (1984) proposed and employed a two-sided filter as a superior alternative to the Kalman filter in the Leser framework, also based on the variance ratio as a smoothing constant. Although it is possible, in principle, to estimate the variance ratios both in the Kalman and the Akaike/Schlicht framework by using maximum-likelihood esti-

Received April 29, 2004. Revised August 16, 2004. Accepted October 15, 2004.
*Department of Economics, University of Munich, Schackstr. 4, 80539 Munich, Germany. Email: schlicht@lmu.de

mates, practical implementation is often not satisfactory, and the rule-of thumb approach prevails.

The aim of this paper is to offer a rather systematic exposition and a straightforward method for estimating the smoothing constant, based on the approach by Schlicht (1984). A maximum likelihood estimator and a related and more intuitive moments estimator will be derived and compared. A simulation study illustrates the performance of these estimators.

The paper is organized as follows. In Section 2 the filter is described; in Section 3, a statistical interpretation of the filter is given that involves some formal parameters. Section 4 gives the estimator for those parameters. In Section 5 it is proved that the descriptive procedure described in Section 2 gives an unbiased maximum-likelihood estimate for the trend, given a smoothing parameter.

Given any smoothing parameter, the covariance matrix of the trend estimate is given in Section 6. Section 7 turns to estimation of the variances by a maximum likelihood method. The variances determine the smoothing parameter. It is shown that the numerical problem can be simplified considerably in several ways.

Section 8 describes a moments estimator for the variances. This estimator is characterized by the property that the computed variances of the error terms are equal to their expectations. In Section 9 it is shown that the likelihood estimates and the moments estimates differ only slightly and approach each other with an increasing length of the time series. This gives intuitive appeal to the maximum likelihood estimator and statistical appeal to the moments estimator.

Section 10 comments on some practical aspects and presents some simulations, and Section 11 offers some concluding comments.

2. The filter

Consider a time series $x \in \mathbb{R}^T$ that is to be decomposed into a trend $y \in \mathbb{R}^T$ and an irregular component $u \in \mathbb{R}^T$:

$$x = y + u. \tag{2.1}$$

Define the trend disturbance $v \in \mathbb{R}^{T-2}$ as

$$v_t = ((y_t - y_{t-1}) - (y_{t-1} - y_{t-2})) \qquad t = 3, 4, \ldots, T$$

or

$$v = Py \tag{2.2}$$

with

$$P := \begin{pmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & & \cdot & \cdot & \cdot \\ 0 & & & 1 & -2 & 1 \end{pmatrix}$$

of order $(T-2) \times T$.

The decomposition of the original series x into trend y and irregular component u is obtained by minimizing the weighted sum of squares

$$u'u + \alpha \cdot v'v = (x-y)'(x-y) + \alpha \cdot y'P'Py$$

with respect to y. This gives the first-order condition

$$(I_T + \alpha \cdot P'P)y = x. \tag{2.3}$$

As $(I + \alpha P'P)$ is positive definite, the second order condition is satisfied in any case and (2.3) can be uniquely solved as

$$y = (I_T + \alpha P'P)^{-1}x. \tag{2.4}$$

Equation (2.4) defines the descriptive filter that associates a trend y to the time series x, depending on the smoothing parameter α.

3. Stochastic interpretation

Equations (2.1) and (2.2) can be embedded in a stochastic model by assuming that the disturbances u and v are normal random variables with variances σ_u^2 and σ_v^2 and zero expectations:

$$u \sim \mathcal{N}(0, \sigma_u^2 I_T), \qquad v \sim \mathcal{N}(0, \sigma_v^2 I_{T-2}). \tag{3.1}$$

This turns x and y into random variables with probability distributions that will be derived in the following. More specifically, the model is defined by (2.1), (2.2) and (3.1). Equation (2.2) describes $\Delta y_t = y_t - y_{t-1}$ as a random walk, and equation (2.1) describes the time series x as the sum of a trend y generated by this random walk and the normal disturbance u. We shall refer to this model as the *true model.*

For purposes of estimation we need a model that explains the observation x as a function of the random variables u and v. This would permit calculating the probability distribution of the observations x contingent on the parameters of the distributions of u and v, *viz.* σ_u^2 and σ_v^2. The true model does not permit such an inference, though, because the matrix P in (2.2) is of rank $T-2$ rather than of rank T. Hence v does not determine a unique y but rather the set of solutions

$$Y := \{P'(PP')^{-1}v + Z\beta \mid \beta \in \mathbb{R}^2\} \tag{3.2}$$

with Z as a $(T \times 2)$-matrix of two orthogonalized solutions $z \in \mathbb{R}^T$ to $Pz = 0$. The matrix Z satisfies

$$PZ = 0, \qquad Z'Z = I_2 \tag{3.3}$$

by definition. For any v we have $y \in Y \Leftrightarrow Py = v$. Equation (2.2) and the set (3.2) give equivalent descriptions of the relationship between y and v in this sense.

In view of (3.2), any solution y to $Py = v$ can be written as

$$y = P'(PP')^{-1}v + Z\beta \tag{3.4}$$

for some $\beta \in \mathbb{R}^2$. As $x = y + u$, equation (2.1) can be re-written as

$$x = u + P'(PP')^{-1}v + Z\beta. \tag{3.5}$$

The model (3.4), (3.5), and (3.1) will be referred to as the *equivalent orthogonally parametrized model*. It implies the true model (2.1), (2.2), and (3.1). Equation (3.4) implies, further, that $\Delta y_t - \Delta y_{t-1} = v_t$. Hence Δy_t is a random walk even though y_t depends, according to (3.4) on past *and* future realizations of v.

Equation (3.5) permits calculation of the density of x dependent upon the parameters of the distributions of u and v and the formal parameters β. In a second step, all these parameters—σ_u^2, σ_v^2, and β—can be determined by the maximum likelihood principle. This will give our maximum likelihood estimates. Our moments estimates—to be introduced later—will build on the equivalent orthogonally parametrized model as well.

The orthogonal parametrization introduced above entails some advantages with respect to symmetry and mathematical transparency, as compared to more usual parameterizations, such as parametrization by initial values. By assuming some initial values $(y_1, y_2)' = c$, the system (2.2) can be solved recursively, giving y as a function of v and c, and the analysis would then proceed in a similar way as indicated above. Theoretically speaking, and with regard to maximum likelihood estimation, all parameterizations are equivalent, but practically initial values are more cumbersome to implement than the formal parameters β. It is for this reason that, in the context of Kalman filtering, initial values are estimated as posterior means, i.e. but by running the filter back and forth in order to determine the necessary initial values iteratively (Akaike (1989), 61-2). The orthogonal parametrization used here will permit us to write down an explicit likelihood function and estimate all relevant parameters in a unified one-shot procedure.

Although the orthogonal parametrization may appear not very intuitive at first sight, it has a straightforward interpretation: The formal parameter vector $\beta \in \mathbb{R}^2$ expresses linear shifts of y that leave the disturbance vector v unaffected. Note that any linear trend $q = (q_1, q_2, \ldots, q_T)'$ with $q_t = a + bt$ for some $(a, b)' \in \mathbb{R}^2$ gives $Pq = 0$. Adding $Z\beta$ means adding such a linear trend to the particular solution $P'(PP')^{-1}v$. This would leave Py unaffected. A possible Z is, for instance, given by

$$z_{t,1} = \frac{1}{\sqrt{T}}, \qquad t = 1, 2, \ldots, T,$$
$$z_{t,2} = \frac{1}{c}\left(t - \frac{T+1}{2}\right)$$

and the constant c chosen such that the orthogonality condition $\sum_t z_{t,2}^2 = 1$ is satisfied. Adjusting β_1 would amount to changing the intercept a of the linear trend $q_t = a+bt$, and changing β_2 would amount to changing the slope b. Adding $Z\beta$ to both x and y would leave the true model (2.1), (2.2) and (3.1) unaffected, and it will turn out that the estimation of the formal parameters β will amount to adjusting the linear part of the estimated trend y to the linear trend found in the observed time series x.

With regard to the relationship between the matrices P and Z, consider

$$\begin{pmatrix} P \\ Z' \end{pmatrix} \begin{pmatrix} P' & Z \end{pmatrix} = \begin{pmatrix} PP' & 0 \\ 0 & I_2 \end{pmatrix}$$

which is of full rank. Inverting both sides, pre-multiplying by (P', Z) and multiplying from the right-hand side by $\binom{P}{Z'}$ implies

$$\begin{pmatrix} P' & Z \end{pmatrix} \begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} P \\ Z' \end{pmatrix} = I_T$$

and hence

$$P'(PP')^{-1}P + ZZ' = I_T. \tag{3.6}$$

The joint distribution of x and y is determined by combining (3.4) and (3.5):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} I_T & P'(PP')^{-1} \\ 0 & P'(PP')^{-1} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} Z \\ Z \end{pmatrix} \beta. \tag{3.7}$$

As the disturbances u and v are independent and normal with variances $\sigma_u^2 I_T$ and $\sigma_v^2 I_{T-2}$, respectively, the vector (u', v') is normal as well:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}(0, S_{uv}).$$

The co-variance matrix is

$$S_{uv} := \begin{pmatrix} \sigma_u^2 \cdot I_T & 0 \\ 0 & \sigma_v^2 \cdot I_{T-2} \end{pmatrix}. \tag{3.8}$$

From (3.7) to (3.8) we obtain

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} Z \\ Z \end{pmatrix} \beta, S_{xy} \right) \tag{3.9}$$

with[1]

$$S_{xy} := \begin{pmatrix} \sigma_u^2 I_T + \sigma_v^2 Q & \sigma_v^2 Q \\ \sigma_v^2 Q & \sigma_v^2 Q \end{pmatrix}$$

and

1 Note that the covariance matrix S_{xy} is not of full rank. Hence (x', y') is distributed on a subspace of $\mathbb{R}^{2T}$ that is determined by the formal parameters β.

$$Q := P'(PP')^{-1}(PP')^{-1}P. \tag{3.10}$$

Note that (3.3) entails

$$Z'Q = 0.$$

From (3.9) we obtain the marginal density of x as

$$x \sim \mathcal{N}(Z\beta, S_x) \tag{3.11}$$

with

$$\begin{aligned} S_x &:= (\sigma_u^2 I_T + \sigma_v^2 Q) \\ &= \sigma_u^2 \left(I_T + \frac{\sigma_v^2}{\sigma_u^2} Q \right). \end{aligned} \tag{3.12}$$

Note further that

$$\begin{aligned} Z' S_x^{-1} &= \frac{1}{\sigma_v^2} Z' \left(I_T - \frac{\sigma_v^2}{\sigma_u^2} Q + \left(\frac{\sigma_v^2}{\sigma_u^2} Q \right)^2 - \left(\frac{\sigma_v^2}{\sigma_u^2} Q \right)^3 + \left(\frac{\sigma_v^2}{\sigma_u^2} Q \right)^4 - \cdots \right) \\ &= \frac{1}{\sigma_v^2} Z'. \end{aligned} \tag{3.13}$$

The marginal density of y for given x is

$$(y \mid x) \sim \mathcal{N}(\bar{y}, S_{y|x})$$

where

$$\bar{y} := Z\beta + \sigma_v^2 Q(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1}(x - Z\beta) \tag{3.14}$$

$$S_{y|x} := \sigma_v^2 Q - \sigma_v^2 Q(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1} \sigma_v^2 Q. \tag{3.15}$$

4. Estimating the formal parameters β

The parameters that need to be estimated are the formal parameter vector β and the variances σ_u^2 and σ_v^2.

The estimation of the formal parameters β is straightforward. Equation (3.11) gives rise to the likelihood function

$$\begin{aligned} L(x, \beta, \sigma_u^2, \sigma_v^2) : = & -\log\det(\sigma_u^2 I_T + \sigma_v^2 Q) \\ & -(x - Z\beta)'(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1}(x - Z\beta). \end{aligned}$$

Maximizing L with respect to β leads to sufficient condition

$$Z'(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1} x = Z'(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1} Z\hat{\beta}.$$

In view of (3.3) and (3.6) this reduces to

$$\hat{\beta} = Z'x. \tag{4.1}$$

5. Estimating the trend y

If we substitute the formal parameters β with the estimator $\hat{\beta}$ in (4.1), we obtain the conditional distribution of the trend y (which is a random variable). It seems sensible to take the expectation of this random variable as our estimator for the trend. This yields

$$\hat{y} := Z\hat{\beta} + Q\left(\frac{\sigma_u^2}{\sigma_v^2} I_T + Q\right)^{-1}(x - Z\hat{\beta}). \tag{5.1}$$

THEOREM 1. *With the smoothing constant α equal to the variance ratio σ_u^2/σ_v^2 the descriptive decomposition* (2.4) *is numerically identical to the estimator* (5.1).

PROOF. Setting $\alpha = \sigma_u^2/\sigma_v^2$ in (5.1) and and ordering terms gives

$$\hat{y} = Z\hat{\beta} + Q(\alpha I_T + Q)^{-1}(x - Z\hat{\beta}). \tag{5.2}$$

Note that

$$Q(\alpha I_T + Q)^{-1} = I_T - \alpha(\alpha I_T + Q)^{-1}$$

which is verified by right-hand multiplication with $(\alpha I_T + Q)$. Inserting this into (5.2) and re-arranging terms gives

$$(\alpha I_T + Q)^{-1}(x - Z\hat{\beta}) = \frac{1}{\alpha}(x - \hat{y}).$$

This can be inserted into (5.2) again, and we obtain

$$\hat{y} = Z\hat{\beta} + Q\frac{1}{\alpha}(x - \hat{y}). \tag{5.3}$$

Pre-multiplication with $\alpha P'P$ yields

$$\alpha P'P\hat{y} = \alpha P'PZ\hat{\beta} + P'PQ(x - \hat{y}). \tag{5.4}$$

As $PZ = 0$, the first term on the right-hand side cancels. From the definition (3.10) of Q and (3.6) it follows that

$$P'PQ = I_T - ZZ'.$$

Substituting this into (5.4) gives

$$\alpha P'P\hat{y} = (x - \hat{y}) - ZZ'(x - \hat{y}). \tag{5.5}$$

Because of (4.1) we have $Z'x = \hat{\beta}$. Pre-multiplying (5.3) by Z' while noting that $Z'Z = I$ and $Z'Q = 0$ results in $Z'\hat{y} = \hat{\beta}$ as well. Hence the last term in (5.5) cancels and we obtain

$$(I_T + \alpha P'P)\hat{y} = x \tag{5.6}$$

which is numerically identical to the normal equation (2.3) that defines the descriptive filter.

Note that pre-multiplying (5.6) by Z' gives $Z'\hat{y} = Z'x$. This is, according to (4.1), the estimate for the formal parameters β which give the linear part of the trend. The estimated trend shares, therefore, its linear component with the original time series.

6. The covariance matrix of the estimates

Consider a given time series x and a realization of the associated trend y. Because x can be viewed as brought about as the sum of the trend y and the disturbance u, we can write:

$$\hat{y} = (I_T + \alpha P'P)^{-1}(y + u). \tag{6.1}$$

Since

$$y = (I_T + \alpha P'P)y - \alpha P'Py$$

and $v = Py$, equation (6.1) can be written as

$$\hat{y} - y = (I_T + \alpha P'P)^{-1}(u - \alpha P'v). \tag{6.2}$$

Equation (6.2) gives the estimation error, and the covariance matrix of this error is calculated as

$$E\{(\hat{y} - y)(\hat{y} - y)'\} = \sigma_u^2(I_T + \alpha P'P)^{-1}. \tag{6.3}$$

For given variances (and therefore a given smoothing constant $\alpha = \sigma_u^2/\sigma_v^2$), equation (6.3) gives the variances of the trend estimates. The square roots of the main diagonal elements of (6.3) give the standard errors of the corresponding point estimates $\hat{y}_t$ of the trend. It is thus possible to guess, for any smoothing parameter α, the precision of the trend estimate.

7. Maximum-likelihood estimation of the variances

In order to estimate the smoothing parameter α, we turn now to estimating the variances σ_u^2 and σ_v^2. A first approach is to simply write down the likelihood function. The distribution of the observations x is given by density function (3.11). Taking logarithms and disregarding constants gives the likelihood

$$\begin{aligned} L(x, \beta, \sigma_u^2, \sigma_v^2) := &-\log\det(\sigma_u^2 I_T + \sigma_v^2 Q) \\ &-(x - Z\beta)'(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1}(x - Z\beta). \end{aligned}$$

By replacing the parameter β with its estimate $\hat{\beta} = Z'x$ from (4.1), we obtain the concentrated likelihood

$$\begin{aligned} L^*(x, \sigma_u^2, \sigma_v^2) := &-\log\det(\sigma_u^2 I_T + \sigma_v^2 Q) \\ &-x'(I_T - ZZ')(\sigma_u^2 I_T + \sigma_v^2 Q)(I_T - ZZ')x. \end{aligned} \tag{7.1}$$

This would suffice, in principle, to estimate the variances σ_u^2 and σ_v^2, but the problem can be simplified considerably. The following theorem states that the likelihood (7.1) can be expressed in terms of the estimated trend $\hat{y}$ and the weighted sum of the variances of the estimates errors $\hat{u}$ and $\hat{v}$ which are defined as follows:

$$\hat{y} := (I_T - \alpha P'P)^{-1}x \tag{7.2}$$

$$\hat{u} := x - \hat{y} \tag{7.3}$$

$$\hat{v} := P\hat{y}. \tag{7.4}$$

Theorem 2. *The likelihood* (7.1) *can be written as*

$$L^*(x, \sigma_u^2, \sigma_v^2) = -\log\det(\sigma_u^2 I_T + \sigma_v^2 Q) - \frac{1}{\sigma_u^2}\hat{u}'\hat{u} - \frac{1}{\sigma_v^2}\hat{v}'\hat{v}. \tag{7.5}$$

Proof. As the first terms of equations (7.1) and (7.5) are identical, it suffices to show that the quadratic forms in these equations are the same. Consider first the quadratic in (7.5). From (7.2) we obtain

$$\hat{u}'\hat{u} = x'(I_T - (I_T + \alpha P'P)^{-1})(I_T - (I_T + \alpha P'P)^{-1})x$$

and

$$\hat{v}'\hat{v} = x'(I_T + \alpha P'P)^{-1}P'P(I_T + \alpha P'P)^{-1}x. \tag{7.6}$$

Because

$$(I_T + \alpha P'P)^{-1} = I_T - \alpha P'P + (\alpha P'P)^2 - (\alpha P'P)^3 + \cdots$$

the matrices $P'P$ and $(I_T + \alpha P'P)^{-1}$ commute and we can re-write equation (7.6) as

$$\hat{v}'\hat{v} = x'P'P(I_T + \alpha P'P)^{-1}(I_T + \alpha P'P)^{-1}x.$$

Combining (7.5) and (7.6) gives

$$\hat{u}'\hat{u} + \alpha\hat{v}'\hat{v} = x'(I_T - (I_T + \alpha P'P)^{-1})x.$$

With

$$A := (I_T - (I_T + \alpha P'P)^{-1}) \tag{7.7}$$

and $\alpha = \sigma_u^2/\sigma_v^2$ the quadratic in equation (7.5) is

$$\frac{1}{\sigma_u^2}\hat{u}'\hat{u} + \frac{1}{\sigma_v^2}\hat{v}'\hat{v} = \frac{1}{\sigma_u^2}x'Ax. \tag{7.8}$$

Consider next the quadratic in (7.1). With

$$B := (I_T - ZZ')'(\alpha I_T + Q)^{-1}(I_T - ZZ') \tag{7.9}$$

it is

$$x'(I_T - ZZ')'(\sigma_u^2 I_T + \sigma_v^2 Q)^{-1}(I_T - ZZ')x = \frac{1}{\sigma_v^2}x'Bx. \tag{7.10}$$

Right-hand multiplication of (7.7) by the non-singular matrices $(I_T + \alpha P'P)$ and $(\alpha I_T + Q)$ and use of (3.6) results in

$$A(I_T + \alpha P'P)(\alpha I_T + Q) = \alpha(I_T + \alpha P'P - ZZ'). \tag{7.11}$$

Equation (7.9) can be re-written as

$$B = (\alpha I_T + Q)^{-1} - \frac{1}{\alpha}ZZ'. \tag{7.12}$$

This makes use of the fact that the matrices $(I_T - ZZ')$ and $(\alpha I_T + Q)^{-1}$ commute, that $(I_T - ZZ')$ is idempotent and that $Z'Q = 0$. Right-hand multiplication of (7.12) by $(\alpha I_T + Q)$ and $(I_T + \alpha P'P)$ yields

$$B(I_T + \alpha P'P)(\alpha I_T + Q) = I_T + \alpha P'P - ZZ'. \tag{7.13}$$

This makes use of the fact that the non-singular matrices $(I_T + \alpha P'P)$ and $(\alpha I_T + Q)$ commute and that $PZ = 0$. Equations (7.11) and (7.13) imply

$$\frac{1}{\sigma_u^2} A = \frac{1}{\sigma_v^2} B.$$

Therefore the expressions given in equations (7.8) and (7.10) are identical.

For purposes of estimation, it is useful to parametrize the likelihood function (7.5) by α and σ_u^2 instead of σ_u^2 an σ_v^2. Because $\sigma_v^2 = \sigma_u^2/\alpha$, we can write:

$$\begin{aligned} L^{**}(x, \sigma_u^2, \alpha) := & -\log\det(\alpha I_T + Q) - \frac{1}{\sigma_u^2}(\hat{u}'\hat{u} + \alpha \hat{v}'\hat{v}) \\ & + T \cdot \log \alpha - T \cdot \log \sigma_u^2. \end{aligned} \tag{7.14}$$

For any given α, the maximization of L^{**} with respect to σ_u^2 leads to the necessary and sufficient conditions

$$\begin{aligned} \frac{\partial L^{**}}{\partial \sigma_u^2} &= -\frac{T}{\sigma_u^2} + \frac{1}{\sigma_u^4}(\hat{u}'\hat{u} + \alpha \hat{v}'\hat{v}) = 0 \\ \left.\frac{\partial^2 L^{**}}{\partial (\sigma_u^2)^2}\right|_{\partial L^{**}/\partial \sigma_u^2 = 0} &= -\frac{T}{\sigma_u^4} < 0 \end{aligned}$$

which imply the estimator

$$\hat{\sigma}_u^2 = \frac{1}{T}(\hat{u}'\hat{u} + \alpha \hat{v}'\hat{v}) \tag{7.15}$$

for the variance of u.

Given any smoothing parameter α, equation (6.3) permits estimating the precision of the trend estimates in terms of the calculated errors:

$$E\{(\hat{y} - y)(\hat{y} - y)'\} = \frac{1}{T}(\hat{u}'\hat{u} + \alpha \hat{v}'\hat{v})(I_T + \alpha \cdot P'P)^{-1}.$$

By inserting (7.15) into (7.14) and disregarding constants, a concentrated likelihood function can be derived that involves the smoothing parameter α as its only parameter:

$$L^{***}(x; \alpha) := -\log\det(\alpha I_T + Q) - T \cdot \log R(\alpha) + T \cdot \log \alpha \tag{7.16}$$

with

$$R(\alpha) := \hat{u}'\hat{u} + \alpha \hat{v}'\hat{v} \tag{7.17}$$

as the weighted sum of estimated squared errors.

With (7.16), maximum likelihood estimation reduces to maximizing over just one parameter. As the solution $\hat{y}$ to the band-diagonal normal equation (2.4) is straightforward, maximization of L^{***} with respect to the smoothing parameter α can be performed numerically. The solution $\hat{y}$ can be calculated for any α. The value of $R(\alpha)$ is calculated *via* (7.2)–(7.4) and (7.17). For any α, the corresponding variances are computed according to (7.15) and $\alpha = \sigma_u^2/\sigma_v^2$ as

$$\hat{\sigma}_u^2 = \frac{1}{T}R(\alpha) \tag{7.18}$$

$$\hat{\sigma}_v^2 = \frac{1}{T}\frac{R(\alpha)}{\alpha}. \tag{7.19}$$

The likelihood function can be further simplified with respect to the first term. Consider

$$(\alpha I_T + Q)(\,P' \;\, Z\,)\begin{pmatrix} P \\ Z' \end{pmatrix} = (\alpha P'P + \alpha ZZ' + I_T - ZZ'). \tag{7.20}$$

As

$$\begin{aligned} \det\left((\,P' \;\, Z\,)\begin{pmatrix} P \\ Z' \end{pmatrix}\right) &= \det\left(\begin{pmatrix} P \\ Z' \end{pmatrix}(\,P' \;\, Z\,)\right) \\ &= \det\begin{pmatrix} PP' & 0 \\ 0 & I_T \end{pmatrix} \\ &= \det(PP'), \end{aligned}$$

equation (7.20) implies

$$\det(PP')\det(\alpha I_T + Q) = \det(I_T + \alpha P'P + (\alpha - 1)ZZ').$$

Right-hand multiplication by $\det(I_T + \alpha P'P)^{-1}$ gives

$$\begin{aligned} &\det(PP')\det(\alpha I_T + Q)\det(I_T + \alpha P'P)^{-1} \\ &\quad = \det(I_T + (\alpha - 1)ZZ'(I_T + \alpha P'P)^{-1}) \\ &\quad = \det(I_T + (\alpha - 1)ZZ'(I_T - \alpha P'P + (\alpha P'P)^2 - (\alpha P'P)^3 + \cdots)) \\ &\quad = \det(I_T + (\alpha - 1)ZZ') \end{aligned}$$

and therefore

$$\frac{\det(\alpha I_T + Q)}{\det(I_T + \alpha P'P)} = \frac{\det(I_T + (\alpha - 1)ZZ')}{\det(PP')}. \tag{7.21}$$

The determinant of $(I_T + (\alpha - 1)ZZ')$ can be evaluated by means of its Eigenvalues. For any symmetric matrix $A \in \mathbb{R}^{n\times n}$, denote the vector of its Eigenvalues by $\Lambda(A)$. The rank $r(A)$ gives the number of non-zero Eigenvalues. The vector of these non-zero Eigenvalues is denoted by $\Lambda^+(A) \in \mathbb{R}^{r(A)}$. The determinant of A is equal to the product of its Eigenvalues.

We have

$$\Lambda(I_T + (\alpha - 1)ZZ') = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + (\alpha - 1)\Lambda(ZZ'). \tag{7.22}$$

Further, $r(Z) = 2$ and ZZ' has rank 2 and two non-zero Eigenvalues of unity.

$$\Lambda^+(ZZ') = \Lambda^+(Z'Z) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In view of (7.22) we conclude that $(I_T + (\alpha - 1)ZZ')$ has $T - 2$ Eigenvalues of one and two Eigenvalues of α. The determinant of $(I_T + (\alpha - 1)ZZ')$ is the product of its Eigenvalues and we can write

$$\det(I_T + (\alpha - 1)ZZ') = \alpha^2.$$

Equation (7.21) reduces thus to

$$\frac{\det(\alpha I_T + Q)}{\det(I_T + \alpha P'P)} = \frac{\alpha^2}{\det(PP')}.$$

Taking logarithms and re-arranging terms yields

$$\log\det(\alpha I_T + Q) = \log\det(I_T + \alpha P'P) + 2\log\alpha - \log\det(PP').$$

Disregarding constants, the likelihood function (7.16) can be written as

$$\mathcal{L}(x;\alpha) = -\log\det(I_T + \alpha P'P) - T \cdot \log R(\alpha) + (T - 2)\log\alpha. \tag{7.23}$$

Note that this formulation avoids, in contrast to the original likelihood function (7.1), the the determination of $Q = P'(PP')^{-1}(PP')^{-1}P$ which is of practical advantage as PP' is large, its inversion is time-consuming, and $(PP')^{-1}$ is *not* sparse. As a consequence, (7.23) requires an evaluation of the determinant of a band-diagonal matrix, while (7.1) would require an evaluation of the the determinant of a *full* matrix at each iteration.

8. A moments estimator for the variances

The likelihood estimation described in the preceding section lacks intuitive appeal, and its small-sample properties are difficult to ascertain. As an alternative, a moments estimator will be devised that is based on the idea that the calculated variances ought to be close to their expectations. (This type of estimator has originally been proposed by Schlicht (1989) in the context of state-space models.) The estimator is derived by equating, at any sample size, the calculated variances with their expectations.

Assume a realization of a trend y (that we can't observe) along with a realization of the time series x (which is taken as a realization of a random variable) for a given set of parameters β, σ_u^2, and σ_v^2. According to (5.6), this gives rise to the estimate $\hat{y}$ as a function of the variance ratio $\alpha = \sigma_u^2/\sigma_v^2$ and of the time series x which is the sum of trend y and disturbance u:

$$\hat{y} = \left(I_T + \alpha\, P'P\right)^{-1}(y+u)\,. \tag{8.1}$$

Since

$$y = (I_T + \alpha P'P)y - \alpha P'Py,$$

and $v = Py$, equation (8.1) can be written as

$$\hat{y} = y + (I_T + \alpha P'P)^{-1}(u - \alpha P'v).$$

Pre-multiplication with P gives

$$\hat{v} = v + P(I_T + \alpha P'P)^{-1}(u - \alpha P'v).$$

In a similar way, from $\hat{u} = x - \hat{y}$ we obtain

$$\hat{u} = u - (I_T + \alpha P'P)^{-1}(u - \alpha P'v).$$

Thus the estimated errors $\hat{u}$ and $\hat{v}$ are linear functions of the the normal random variables u and v:

$$\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \begin{pmatrix}(I_T - M) & \alpha M P' \\ PM & I_{T-2} - \alpha PMP'\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}$$

$$\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \left(I_{2T-2} - \begin{pmatrix}I_T & 0\\ 0 & P\end{pmatrix}\begin{pmatrix}M & -M\\ -M & M\end{pmatrix}\begin{pmatrix}I_T & 0\\ 0 & \alpha P'\end{pmatrix}\right)\begin{pmatrix}u\\ v\end{pmatrix}$$

with

$$M := (I_T + \alpha P'P)^{-1}$$

and their joint distribution can be calculated:

$$\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \mathcal{N}\left(0, \begin{pmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{pmatrix}\right)$$

with

$$\begin{aligned}
S_{11} &:= \sigma_u^2(I_T - M)^2 + \sigma_v^2\alpha^2 MP'PM\\
S_{12} &:= \sigma_u^2(I_T - M)MP' + \sigma_v^2\alpha MP'(I_{T-2} - \alpha PMP')\\
S_{21} &:= \sigma_u^2 PM(I_T - M) + \sigma_v^2\alpha(I_{T-2} - \alpha PMP')PM\\
S_{22} &:= \sigma_u^2 PM^2P' + \sigma_v^2(I_{T-2} - \alpha PMP')^2.
\end{aligned}$$

From this, the expectation of the average squared errors can be determined:

$$E\{\hat{u}'\hat{u}\} = \sigma_u^2 \cdot \operatorname{tr}(I_T - M)^2 + \sigma_u^2 \alpha \operatorname{tr}(M^2 P' P) \tag{8.2}$$

$$E\{\hat{v}'\hat{v}\} = \sigma_u^2 \cdot \operatorname{tr}(P M^2 P') + \sigma_v^2 \cdot \operatorname{tr}(I_{T-2} - \alpha P M P')^2. \tag{8.3}$$

Note that

$$\begin{aligned} \operatorname{tr}(I_T - M)^2 + \alpha \operatorname{tr}(M^2 P' P) &= \operatorname{tr}(I_T - 2M + (I_T + \alpha P' P) M^2) \\ &= \operatorname{tr}(I_T - M) \\ &= T - \operatorname{tr}(M) \end{aligned}$$

and

$$\begin{aligned} &\alpha \operatorname{tr}(P M^2 P') + \operatorname{tr}(I_{T-2} - \alpha P M P')^2 \\ &= \operatorname{tr}(\alpha P M^2 P' + I_{T-2} - 2\alpha P M P' + \alpha^2 (P M P')^2) \\ &= \operatorname{tr}(\alpha P M (I_T + \alpha P' P) M P' + I_{T-2} - 2\alpha P M P') \\ &= \operatorname{tr}(I_{T-2} - \alpha P M P') \\ &= (T-2) - \operatorname{tr}(\alpha M P' P). \end{aligned} \tag{8.4}$$

Because

$$M(I_T + \alpha P' P) = I_T$$

we have

$$\operatorname{tr}(\alpha M P' P) = T - \operatorname{tr}(M).$$

Inserting this into (8.4) gives

$$\operatorname{tr}(I_T - M)^2 + \alpha \operatorname{tr}(M^2 P' P) = \operatorname{tr}(M) - 2$$

and (8.2)–(8.3) reduce to

$$\begin{aligned} E\{\hat{u}'\hat{u}\} &= \sigma_u^2 (T - \operatorname{tr}(M)) \\ E\{\hat{v}'\hat{v}\} &= \sigma_v^2 (\operatorname{tr}(M) - 2). \end{aligned}$$

The moments estimators for the variances, denoted by $\breve{\sigma}_u^2$ and $\breve{\sigma}_v^2$, are obtained by equalizing the estimated moments $\hat{u}'\hat{u}$ and $\hat{v}'\hat{v}$ with their expectations:

$$\hat{u}'\hat{u} = \breve{\sigma}_u^2 (T - \operatorname{tr}(M)) \tag{8.5}$$

$$\hat{v}'\hat{v} = \breve{\sigma}_v^2 (\operatorname{tr}(M) - 2). \tag{8.6}$$

Note that the estimated moments $\hat{u}'\hat{u}$ and $\hat{v}'\hat{v}$, as implied by (7.2)–(7.4) are functions of the observations x and the variance ratio $\breve{\alpha} = \breve{\sigma}_u^2 / \breve{\sigma}_v^2$ and, thus, of the variances $\breve{\sigma}_u^2$ and $\breve{\sigma}_v^2$, and that the matrix M depends on the variance ratio as well. Hence the solution to (8.5)–(8.6) amounts to finding a fix-point.

The system can be written equivalently as

$$\breve{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{T - \operatorname{tr}(M)} \tag{8.7}$$

$$\breve{\sigma}_v^2 = \frac{\hat{v}'\hat{v}}{\operatorname{tr}(M) - 2}. \tag{8.8}$$

One way of estimating the variances is, thus, to find a fix-point of (8.5)–(8.6) or (8.7)–(8.8). Another way is the following.

Consider the function

$$\mathcal{H}(x,\alpha) = -\log\det(I+\alpha P'P) - (T-2)\cdot\log R(\alpha) + (T-2)\cdot\log\alpha. \tag{8.9}$$

The following theorem states that the moments estimator can be derived by maximizing the function $\mathcal{H}(x,\alpha)$.

THEOREM 3. *The moments estimators, as defined by equations* (8.7) *and* (8.8), *can be obtained by maximizing the function* $\mathcal{H}(x,\alpha)$ *defined in* (8.9) *with respect to* α. *The variances are computed from the maximizing value* $\tilde{\alpha}$ *as*

$$\check{\sigma}_u^2 = \frac{1}{T-2}R(\tilde{\alpha}) \tag{8.10}$$

$$\check{\sigma}_v^2 = \frac{1}{T-2}\frac{R(\tilde{\alpha})}{\tilde{\alpha}}. \tag{8.11}$$

PROOF. With $M=(I_T+\alpha P'P)^{-1}$ we have

$$\begin{aligned}\frac{d\log\det(I_T+\alpha P'P)}{d\alpha} &= \frac{d}{d\alpha}\left(T\log\alpha+\log\det\left(\frac{1}{\alpha}I_T+P'P\right)\right)\\ &= \frac{1}{\alpha}(T-\mathrm{tr}(M)).\end{aligned}$$

Note further that

$$\begin{aligned}R(\alpha) &= x'(I_T-M)^2x+\alpha x'MP'PMx\\ &= x'(I_T-2M+M^2+\alpha MP'PM)x\\ &= x'(I_T-2M+M(I_T+\alpha P'P)M)\\ &= x'(I_T-M)x\\ &= x'x-x'Mx.\end{aligned} \tag{8.12}$$

Consider

$$\begin{aligned}\frac{\partial M}{\partial\alpha} &= \frac{\partial(\alpha(\frac{1}{\alpha}I_T+P'P))^{-1}}{\partial\alpha}\\ &= -MP'PM.\end{aligned} \tag{8.13}$$

From (8.12) and (8.13) we obtain

$$\begin{aligned}R'(\alpha) &= x'MP'PMx\\ &= \hat{v}'\hat{v}.\end{aligned} \tag{8.14}$$

Using these results, the derivative of $\mathcal{H}(x,\alpha)$ with respect to α is calculated as

$$\frac{\partial\mathcal{H}}{\partial\alpha} = -\frac{1}{\alpha}(T-\mathrm{tr}(M)) - (T-2)\frac{\hat{v}'\hat{v}}{R}+\frac{(T-2)}{\alpha}.$$

Putting this to zero yields

$$\tilde{\alpha} = \frac{\operatorname{tr}(M) - 2}{T - \operatorname{tr}(M)} \cdot \frac{\hat{u}'\hat{u}}{\hat{v}'\hat{v}}. \tag{8.15}$$

Evaluating the right-hand side of (8.10) by using (8.15) gives

$$\begin{aligned} \frac{1}{T-2} R(\tilde{\alpha}) &= \frac{1}{T-2}\left(\hat{u}'\hat{u} + \frac{\operatorname{tr}(M)-2}{T-\operatorname{tr}(M)} \cdot \frac{\hat{u}'\hat{u}}{\hat{v}'\hat{v}}\hat{v}'\hat{v}\right) \\ &= \frac{\hat{u}'\hat{u}}{T - \operatorname{tr}(M)} \end{aligned}$$

which is identical to (8.10). Evaluating the right-hand side of (8.11) gives

$$\begin{aligned} \frac{1}{T-2}\frac{R(\tilde{\alpha})}{\tilde{\alpha}} &= \frac{1}{T-2}\left(\hat{u}'\hat{u} + \frac{\operatorname{tr}(M)-2}{T-\operatorname{tr}(M)} \cdot \frac{\hat{u}'\hat{u}}{\hat{v}'\hat{v}}\hat{v}'\hat{v}\right) \frac{T-\operatorname{tr}(M)}{\operatorname{tr}(M)-2} \cdot \frac{\hat{v}'\hat{v}}{\hat{u}'\hat{u}} \\ &= \frac{\hat{v}'\hat{v}}{\operatorname{tr}(M) - 2} \end{aligned}$$

which is identical to (8.11).

9. The relationship between the maximum-likelihood and the moments estimator

With the aid of the function $\mathcal{H}$, the relationship between the maximum-likelihood estimator and the moments estimator can be gauged easily.

Note that

$$\mathcal{H}(x, \alpha) - \mathcal{L}(x; \alpha) = 2R(x, \alpha)$$

and denote, for some given x, the maximizer of $\mathcal{L}(x; \alpha)$ by $\hat{\alpha}$ and the maximizer of $\mathcal{H}(x; \alpha)$ by $\check{\alpha}$. By definition we have

$$\begin{aligned} \mathcal{L}(x; \hat{\alpha}) &\geq \mathcal{L}(x; \check{\alpha}) \\ \mathcal{H}(x; \check{\alpha}) &\geq \mathcal{H}(x; \hat{\alpha}). \end{aligned}$$

This implies

$$\begin{aligned} &\mathcal{H}(x; \check{\alpha}) - \mathcal{L}(x; \check{\alpha}) \geq \mathcal{H}(x; \hat{\alpha}) - \mathcal{L}(x; \hat{\alpha}) \\ &R(\check{\alpha}) \geq R(\hat{\alpha}) \end{aligned}$$

with strict inequality if the maximizers $\check{\alpha}$ and $\hat{\alpha}$ are unique and distinct. As R is, according to (8.14), increasing in α, the method of moments will yield a larger smoothing constant than the maximum-likelihood method.

With regard to variance estimation we have from (7.18) and (8.10)

$$\hat{\sigma}_u^2 = \frac{1}{T} R(\hat{\alpha}), \qquad \check{\sigma}_u^2 = \frac{1}{T-2} R(\check{\alpha})$$

which implies that the moments estimate $\check{\sigma}_u^2$ is larger than the maximum likelihood estimate $\hat{\sigma}_u^2$. From (7.19) and (8.11) we have further

$$\hat{\sigma}_v^2 = \frac{1}{T}\frac{R(\hat{\alpha})}{\hat{\alpha}}, \qquad \check{\sigma}_v^2 = \frac{1}{T-2}\frac{R(\check{\alpha})}{\check{\alpha}}.$$

Equations (7.17) and (8.14) imply $\frac{R'}{R} < \frac{1}{\alpha}$. Hence $\frac{R}{\alpha}$ is decreasing in α. For large T this effect will dominate and we will have $\hat{\sigma}_v^2 > \check{\sigma}_v^2$, but for small T the reverse effect may come about.

As $\frac{T-2}{T} \to 1$ for $T \to \infty$, the moments criterion

$$-\frac{1}{T}\log\det(I + \alpha P'P) - \frac{T-2}{T}\log R(\alpha) + \frac{T-2}{T}\log\alpha \tag{9.1}$$

and the likelihood criterion

$$-\frac{1}{T}\log\det(I + \alpha P'P) - \log R(\alpha) + \frac{T-2}{T}\log\alpha \tag{9.2}$$

become identical for large T and the estimates will coincide. The same holds true for the estimates of the variances (7.18), (7.19) and (8.10), (8.11).

10. Notes on numerical performance

A practical example is provided in Figure 1[2]. The method of moments gives $\sigma_u^2 = 4.95E{-}5$, $\sigma_v^2 = 1.71E{-}5$, and $\alpha = 2.89$. The maximum likelihood estimates are $\sigma_u^2 = 4.35E-5$, $\sigma_v^2 = 2.10E-5$, and $\alpha = 2.06$. The corresponding graphs are practically indistinguishable, the maximum relative difference between the two estimates being below 2 percent. Only rather drastic changes in the smoothing constant—like increasing or decreasing it by a factor of ten—produce significant changes (Figure 2).

These observations do not tell much, however, about how well the method recovers the smoothing constant and the variances of the time series. Some simulations were conducted in order to obtain an impression about this aspect of performance, and also to compare the two variants of the method, *viz.* maximum likelihood and method of moments. It is beyond the scope of this paper to present a full-fledged Monte-Carlo study. The following remarks are intended to just convey an overall impression.

For the simulations done, the method works reasonably well in both variants. Consider the estimation of the smoothing constant first, or rather its $\log_{10}(\alpha)$, because this seems to be the more relevant quantity. Figure 3 depicts the frequency distribution for the estimates of the smoothing constant that are obtained by generating 1000 random series according to equations (2.1), (2.2), and (3.1) with variances $\sigma_u^2 = 10$ and $\sigma_v^2 = 1$ (corresponding to a smoothing constant $\log_{10}(\alpha) = 1$) for alternative lengths $T = 30$, $T = 60$, and $T = 120$, respectively and using the method of moments estimator.

As expected, the estimates are less reliable for short series and more reliable for long series.

2 All computations are made using the Mathematica Package by Ludsteck (2004).

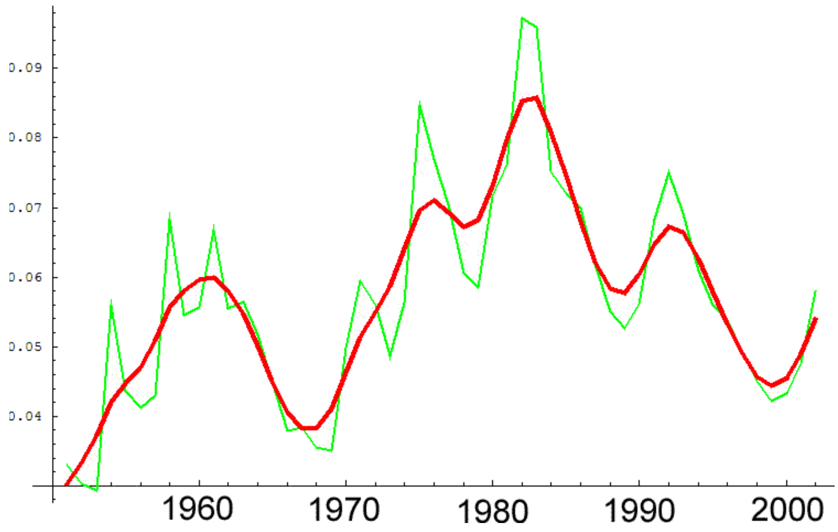

Figure 1. US unemployment 1951–2002, source: US Department of Commerce, Bureau of Labor Statistics. Estimated parameters: $\sigma_u^2 = 4.95E{-}5$, $\sigma_v^2 = 1.71E{-}5$, and $\alpha = 2.89$.

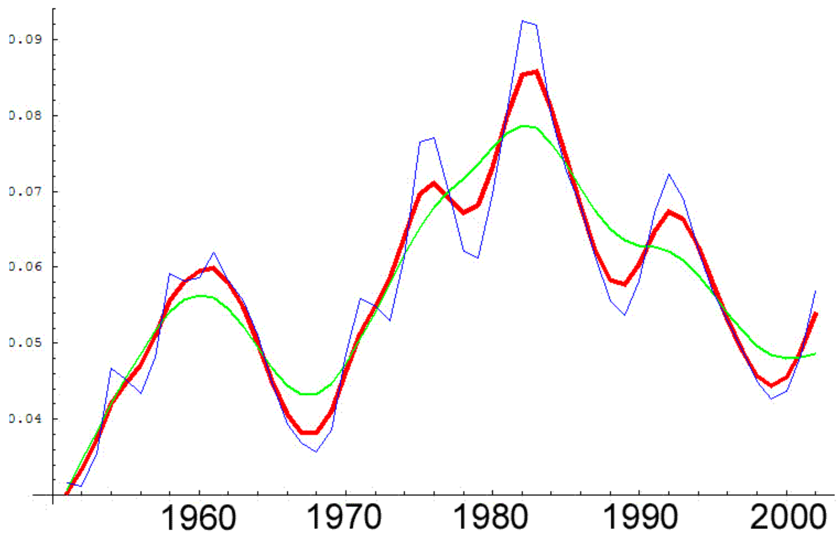

Figure 2. Effect of the smoothing constant: Heavily drawn curve: smoothed with estimated value, smooth curve: smoothed with tenfold of estimated value, thin curve: smoothed with a tenth of estimated value.

For practical purposes, and regarding the smoothing constant, the maximum-likelihood estimates and the moments estimates are nearly identical (Table 1). There may be a slight but certainly insignificant advantage for the maximum-likelihood estimator regarding the standard deviation and a slight advantage for the moments estimator regarding numerical stability (fewer crashes under the

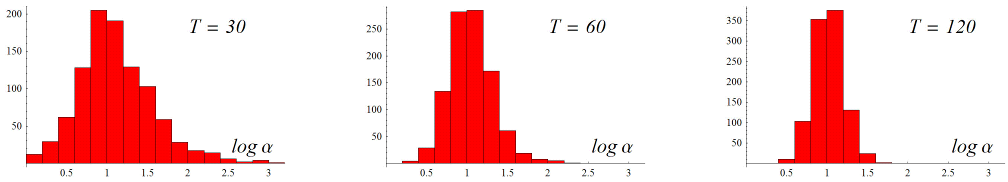

Figure 3. Distribution of estimates of $\log_{10}(\alpha)$ for time series of different length with $\sigma_u^2 = 10$ and $\sigma_v^2 = 1$, 1000 trials each.

Table 1. Estimation of the smoothing constant: Performance of the maximum likelihood (ML) and the moments estimator (MM) for time series of different length. True variances are $\sigma_u^2 = 10$ and $\sigma_v^2 = 1$, reported statistics based on 1000 successful trials each.

Length	Estimator	Mean	Median	Stdev	Min	Max	Crashes
15	ML	1.00	0.92	0.74	−0.62	7.66	272
	MM	1.05	0.96	0.76	−1.57	7.82	241
30	ML	1.40	0.96	0.53	−0.49	4.29	16
	MM	1.08	1.01	0.50	−0.17	4.63	15
60	ML	1.01	0.99	0.28	0.12	2.30	0
	MM	1.05	1.03	0.28	0.25	2.25	0
120	ML	1.01	0.99	0.18	0.45	1.66	0
	MM	1.02	1.01	0.18	0.48	1.77	0
240	ML	1.00	1.00	0.13	0.60	1.53	0
	MM	1.1	1.1	0.14	0.57	1.46	0

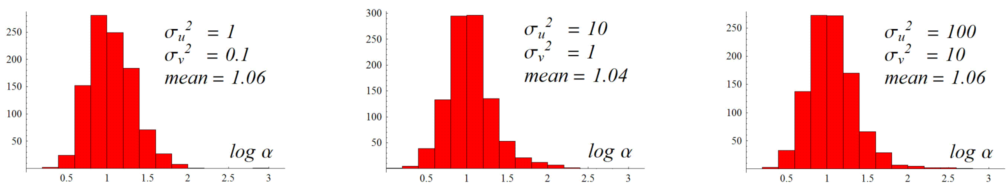

Figure 4. Changing variances while keeping the variance ratios constant does not affect the result. (Standard deviations are in the range 0.28–0.30 in all three cases, $T = 60$, 1000 trials each.)

same algorithm).

The decomposition depends on the smoothing constant, *viz.* the ratio of the variances, rather than the absolute magnitude of the variances which are affected by scaling. This independence is reconfirmed in the simulations (Figure 4).

Finally, Figure 5 gives the results when variance ratios are changed by a factor of 10. This shifts the distribution on the logarithmic scale to the left or to the right by one unit.

Table 2 gives the results for the smoothing constant and the variances. There,

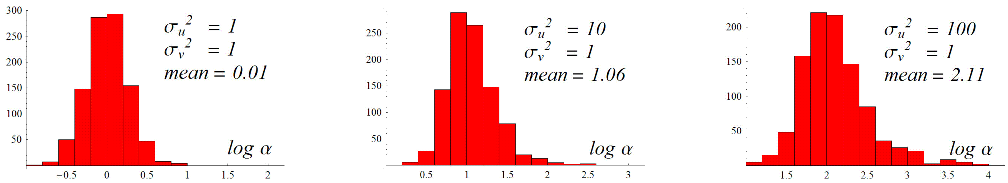

Figure 5. Increasing the variance ratio by some factor shifts the distribution of estimates to the right by the same factor. The true means of 0, 1, and 2 are nicely recovered. The standard deviation increases from 0.26 to 0.30 to 0.43 with an increasing variance ratio. ($T = 60$, 1000 trials each.)

Table 2. Performance of the maximum likelihood (ML) and the moments estimator (MM) regarding the estimation of α, σ_u^2, and σ_v^2 for time series of different length. True values are $\log_{10}\alpha = 1$, $\log_{10}\sigma_u^2 = 1$ and $\log_{10}\sigma_v^2 = 0$; standard deviations in parentheses, reported statistics based on 5000 successful trials each.

Length	Estimator	$\log_{10}\alpha$	$\log_{10}\sigma_u^2$	$\log_{10}\sigma_v^2$
30	ML	1.06	0.93	−0.13
		(0.54)	(0.15)	(0.48)
	MM	1.14	0.98	−0.16
		(0.53)	(0.14)	(0.48)
60	ML	1.02	0.97	−0.05
		(0.30)	(0.10)	(0.26)
	MM	1.05	0.99	−0.06
		(0.29)	(0.10)	(0.26)

is, again, no significant difference between the variants noticeable.

In conclusion, there is no big difference between the maximum likelihood estimator and the moments estimator even in short time series.

11. Notes on modelling

The trend filter discussed here has given rise to two strands of thought. One, originally proposed by Akaike (1980) and Schlicht (1984) and also alluded to by Hodrick and Prescott (1997), relates to state-space modelling; the other, starting with King and Rebelo (1993), looks at performance in the frequency domain. The present paper falls into the first category and implies some scepticism concerning the second.

The state-space literature tends to rely on Kalman filtering. As Kalman filters are one-sided filters, they are never efficient in the sense of using all available information for estimating trend values at intermediate points in time. The filter proposed here is, in contrast, two-sided and uses all information available. Further, the orthogonal parametrization avoids the problem of estimating initial values and allows for a unified and mathematically more transparent treatment of the maximum likelihood estimator as well as the moments estimator than can be achieved by parametrization in terms of initial values.

It turns out that, in this setting, the maximum-likelihood estimator and the moments estimator are practically identical. Hence the intuitive interpretation suggested by the moments estimator carries over to the maximum-likelihood estimator, and the statistical appeal of the maximum-likelihood estimator carries over to the moments estimator.

From the perspective taken in this paper, the frequency interpretation is problematic, as the smoothing constant is unrelated to any frequency found in the trend. Further, two time series generated by the same trend but affected by different disturbances u would require different smoothing constants for optimal recovery.

Yet the filter is not well suited for determining the trend as distinct from the business cycle. This would require an integrated full-fledged method for seasonal *and* cyclical adjustment.

Acknowledgements

The author thanks Johannes Ludsteck for several helpful suggestions and for providing the Mathematica Package (Ludsteck (2004)) used for the computations in this paper. The comments of two anonymous referees of this journal were extremely helpful and prompted important improvements.

References

Akaike, H. (1980). Seasonal adjustment by Bayesian modeling, *Journal of Time Series Analysis*, **1**(1), 1–13.

Akaike, H. (1989). Bayesian modeling for time series analysis, *Advances in Statistical Analysis and Statistical Computing* (eds. R. S. Mariano, K. S. Lim and K. D. Ling), pp. 59–69, Greenwich (Conn.) and London, JAI Press.

Henderson, R. (1924). A new method of graduation, *Transactions of the Actuarial Society of America*, **25**, 29–40.

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

King, R. G. and Rebelo, S. T. (1993). Low frequency filtering and real business cycles, *Journal of Economic Dynamics and Control*, **17**, 207–231.

Leser, C. E. V. (1961). A simple method of trend construction, *Journal of the Royal Statistical Society. Series B (Methodological)*, **23**, 91–107.

Ludsteck, J. (2004). *HPFilter*. Mathematica Package. Wolfram Research, online at —http://library.wolfram.com/infocenter/MathSource/5161/—.

Schlicht, E. (1984). Seasonal adjustment in a stochastic model, *Statistische Hefte*, **25**, 1–12. Online at —http://www.lrz.de/˜ekkehart—.

Schlicht, E. (1989). *Variance Estimation in a Random Coefficients Model*, Paper presented at the Econometric Society European Meeting Munich. Online at —http://www.lrz.de/˜ekkehart—.

Whittaker, E. T. (1923). On a new method of graduation, *Proceedings of the Edinburgh Mathematical Society*, **41**, 63–75.