An optimality criterium for the multivariate Hodrick-Prescott filter

Azzouz Dermoune[*], Boualem Djehiche[†] and Nadji Rahmania [‡]

June 17, 2008

Abstract

The univariate Hodrick-Prescott filter depends on the noise-to-signal ratio that acts as a smoothing parameter. We first propose an optimality criterium for choosing the best smoothing parameters, and show that the noise-to-signal ratio is the unique minimizer of this criterium. We then propose a multivariate extension of the filter and show that there is a whole class of positive definite matrices that satisfy a similar optimality criterium.

JEL classifications: E32, C22.

AMS 1991 subject classifications: 62M10, 62G15.

Key words and phrases: Adaptive estimation, Gaussian process, Hodrick-Prescott filter, noise-to-signal ratio, orthogonal parametrization.

[*]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

1 Introduction

The univariate Hodrick-Prescott filter (HP in short) was first introduced in actuarial science to estimate trends from claims data and, since the appearence of the seminal paper by Hodrick and Prescott (1980) it is now widely used in economics and finance to estimate and predict business cycles and trends in financial data series. The filter as described in Hodrick and Prescott (1980) and (1997) defines 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 α, called the smoothing parameter. Assuming the components of the residual (noise) $u = x - y$ independent and Gaussian with the same variance as well as the components of the signal v defined by $v_t = y_{t+2} - 2y_{t+1} + y_t$, which amounts to considering a Gaussian random walk model of the trend, Hodrick and Prescott (1980) (see also Schlicht (2006)) suggested that the best smoothing parameter is the positive parameter α for which the trend $\hat{y}(\alpha, x)$ that minimizes the above weighted sum is *numerically* equal to the best predictor, i.e. the conditional expectation $E[\, y |\, x]$, of y given the time series x. As shown in Schlicht (2006), Theorem 1, the so-called noise-to-signal ratio i.e. the ratio of the variance of the noise u to the variance of the signal v, is the only parameter that satisfies this criterium.

In the first part of this study, we propose to choose the best smoothing parameters for the univariate HP filter which minimize the gap (using the Euclidean norm) between $\hat{y}(\alpha, x)$ and $E[\, y |\, x]$,

$$\alpha^* = \arg\min_{\alpha > 0} \| E[\, y |\, x] - \hat{y}(\alpha, x) \|^2, \quad \text{for all realisations of } x, \tag{1.1}$$

and show that the noise-to-signal ratio is the unique optimal solution.

Razzak and Dennis (1995) were first to suggest a multivariate version of the HP filter for which a trend minimizes the following weighted sum $\sum_{t=1}^{T}(x_t - y_t)^2 + \sum_{t=1}^{T-2} \alpha_t (y_{t+2} - 2y_{t+1} + y_t)^2$, where, the smoothing parameter is instead a diagonal matrix $\alpha = \text{diag}\{\alpha_1, \ldots, \alpha_T\}$, with, for each $t = 1, \ldots, T$, α_t is assumed to be the ratio of the variance of the noise u_t to the variance of the signal v_t. For other models, we refere to Reeves *et al.* (2000) and the references therein.

In the second part of this study, we also propose a multivariate extension of the HP filter, where, for instance, the smoothing parameter is a pair of positive definite matrices, and show that there is a whole family of positive definte matrices that are optimal for a similar optimality criterium to (1.1). In fact, this family of optimal parameters is an *equivalence class*, since these parameters give the same value of the trend $\hat{y}$. In the particular model of Razzak and Dennis (1995), we show that the noise-to-signal ratio i.e. the ratio of the variance of the noise u_t to the variance of the signal v_t is optimal but far from being unique.

The paper is organized as follows. In Section 2, we establish an optimality crtierium for the best smoothing parameters of the univariate HP filter. In Section 3, we propose a multivariate version of the HP filter and an optimality criterium for choosing the best smoothing parameters. It turns out to that the optimal solutions is a whole set of pairs of positive definite matrices that satisfy a balance equation involving the covariance matrices of both the noise and the signal.

2 Univariate Hodrick-Prescott filter

Let $x = (x_1, ..., x_T) \in \mathbb{R}^T$ 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) $u \in \mathbb{R}^T$:

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

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

$$\|x - y\|^2 + \alpha\|D^2 y\|^2 \tag{2.2}$$

with respect to y, where for $a \in \mathbb{R}^T$, $\|a\|^2 = \sum_{i=1}^T a_i^2$. Here, $D^2 y$ is the trend disturbance obtained by acting the second order forward shift operator D^2 on the trend $y = (y_1, y_2, \dots, y_T)$:

$$D^2 y_t := (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t), \qquad t = 1, 2, \dots, T-2,$$

or, equivalently,

$$D^2 y_t := 2\left(\frac{y_{t+2} + y_t}{2} - y_{t+1}\right), \qquad t = 1, 2, \dots, T-2,$$

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,

$$Py(t) = D^2 y_t, \qquad t = 1, \dots, T-2, \tag{2.3}$$

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

$$P := \begin{pmatrix} 1 & -2 & 1 & \dots & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ & \dots & \dots & \dots & \dots & \\ 0 & \dots & \dots & 1 & -2 & 1 \end{pmatrix}. \tag{2.4}$$

The first term in (2.2) measures a goodness-of-fit by minimizing the 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 the deviation between the trend value y_{t+1} and the linear interpolation between y_t and y_{t+2}.

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 (2006) for further details)

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

where the $T \times 2$-matrix Z satisfies

$$PZ = 0, \qquad Z'Z = I_2, \tag{2.5}$$

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

$$x = u + P'(PP')^{-1}v + Z\gamma, \tag{2.6}$$

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

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

$$y(\alpha, x) = (I_T + \alpha P'P)^{-1}x, \tag{2.7}$$

where I_T denotes the $T \times T$ identity matrix. Eq. (2.7) defines the descriptive filter that associates a trend y to the time series x, depending on the smoothing parameter α and the disturbance operator P.

2.1 A criterium for choosing the best smoothing parameter

Following Hodrick and Prescott (1980) and Schlicht (2006), a way to estimate the smoothing parameter α is to let the optimal solution $y(\alpha, x)$ in (2.7) be the best predictor of y given the time series x, i.e.

$$y(\alpha, x) \approx E[\, y|\, x]. \tag{2.8}$$

We will give a precise meaning of this relation, and show that the unique solution α^* of (2.8) is given by

$$\alpha^* = \arg\min_{\alpha > 0} \|E[\, y|\, x] - y(\alpha, x)\|^2. \tag{2.9}$$

Both approaches of estimating α assume that we are able to compute explicitly this conditional expectation, which is not always the case. The Gaussian 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 (2.6), this can be achieved through imposing a model for the joint distribution of (u, v).

In the literature (cf. e.g. Hodrick and Prescott (1997) and Schlicht (2006)), 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 α, using (2.7) and (2.8), feasible. In particular, as suggested in Hodrick and Prescott (1997), assuming further that the noise term u 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:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}(0, \Sigma_{uv}), \tag{2.10}$$

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 (2.6), (x, y) is normally distributed:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} Z \\ Z \end{pmatrix}\gamma, \Sigma_{xy}\right), \tag{2.11}$$

with covariance matrix

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

where,

$$Q := P'(PP')^{-1}(PP')^{-1}P$$

is a symmetric matrix.

Recall the following properties relating P, Z and Q defined above

$$P'\left(PP'\right)^{-1}P + ZZ' = I_T,\ QZ = 0, \quad Z'Q = 0,\ Z'Z = I_2,\ \operatorname{trace}(ZZ') = 2 \tag{2.12}$$

and

$$Q[\frac{\sigma_u^2}{\sigma_v^2} I_T + Q]^{-1} = [\frac{\sigma_u^2}{\sigma_v^2} I_T + Q]^{-1} Q. \tag{2.13}$$

This yields that

$$Q[\frac{\sigma_u^2}{\sigma_v^2} I_T + Q]^{-1} Z = 0, \tag{2.14}$$

and

$$y(\alpha, x) = (I_T + \alpha P'P)^{-1}x = ZZ'x + Q[\alpha I_T + Q]^{-1}x. \tag{2.15}$$

Moreover, the maximum likelihood estimator of γ is explicitly given as follows:

$$\arg\min_{\gamma}(x - Z\gamma)'[\sigma_u^2 I_T + \sigma_v^2 Q]^{-1}(x - Z\gamma) = Z'x.$$

By (2.11), an explicit expression of the conditional expectation of the trend y given the time series x reads:

$$E[\,y|\,x] = Z\gamma + \sigma_v^2 Q\left[\sigma_u^2 I_T + \sigma_v^2 Q\right]^{-1}(x - Z\gamma). \tag{2.16}$$

Now, Criteruim (2.8) can be understood in the following sense: The best smoothing parameter is the positive α for which the following identity holds.

$$E(y \,|\, x) - Z\gamma = y(\alpha, x) - Z\hat{\gamma} \text{ for all realizations of } \ x \text{ in } \mathbb{R}^T, \tag{2.17}$$

or, equivalently,

$$Z\gamma + \sigma_v^2 Q \left[\sigma_u^2 I_T + \sigma_v^2 Q\right]^{-1} (x - Z\gamma) = (I_T + \alpha P'P)^{-1} x - ZZ'x \tag{2.18}$$

Thanks to (2.15), Equation (2.18) holds if and only if $\alpha = \sigma_u^2/\sigma_v^2$. This gives a characterization of the parameter α as the noise-to-signal ratio, $\alpha = \sigma_u^2/\sigma_v^2$.

2.2 Optimality of the noise-to-signal ratio

Using the criterion (2.9), we have the following result.

Proposition 2.1 *We have*

$$\alpha^* = \sigma_u^2/\sigma_v^2 = \arg\min_{\alpha>0} \left\| E[\, y|\, x] - y(\alpha, x) \right\|^2. \tag{2.19}$$

Moreover, the error (optimal gap)

$$E[\, y|\, x] - y(\alpha^*, x) = Z\,(\gamma - Z'x)$$

is a centered Gaussian vector with covariance matrix

$$cov\,(Z\,(Z'x - \gamma)) = \sigma_u^2 ZZ'.$$

In particular,

$$E[\|E[\, y|\, x] - y(\alpha^*, x)\|^2] = E[\|Z\,(\gamma - Z'x)\,\|^2] = \sigma_u^2\, trace(ZZ') = 2\sigma_u^2.$$

Proof. Recall Identity (2.15):

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

Therefore, using (2.16), we have

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

Now, since $Z'Q = 0$, we also have

$$(Z\,(\gamma - Z'x))'\, Q\left\{\left[\sigma_u^2/\sigma_v^2 I_T + Q\right]^{-1} - [\alpha I_T + Q]^{-1}\right\} x = 0.$$

Hence,

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

This yields

$$\|E[y|x] - y(\alpha, x)\|^2 \geq \left\|E[y|x] - y(\sigma_u^2/\sigma_v^2, x)\right\|^2 = \|Z(\gamma - Z'x)\|^2,$$

for all $\alpha > 0$. Thus $\alpha^* = \sigma_u^2/\sigma_v^2$ is optimal.

The covariance matrix of the centered Gaussian vector $Z(Z'x - \gamma)$ is

$$cov(Z(Z'x - \gamma)) = E[Z(Z'x - \gamma)(x'Z - \gamma')Z'] = ZZ'E[(x - Z\gamma)(x - Z\gamma)']ZZ'.$$

In particular,

$$E[\|Z(\gamma - Z'x)\|^2] = \sigma_u^2 \text{trace}(ZZ') = 2\sigma_u^2.$$

□

3 A multivariate Hodrick-Prescott filter

A natural extension of the univariate HP-filter is to be able to perform a similar trend detection for a multidimensional time series of observables. Let x be a d-dimensional time series of observations at T instants, that we may represent as a (column) vector in $\mathbb{R}^{dT}$:

$$x = ((x_1^1, \ldots, x_T^1), \ldots, (x_1^d, \ldots, x_T^d)) \in \mathbb{R}^{dT}.$$

We would like to decompose x into a nonstationary trend $y \in \mathbb{R}^{dT}$ and a cyclical residual component (noise term) $u \in \mathbb{R}^{dT}$:

$$x = y + u. \tag{3.1}$$

by minimizing a weighted sum of squares similar to (2.2), with an appropriate combination of matrices that act as a smoothing parameter similar to α, given a trend-smoothing operator A similar to the shift operator P in the standard HP-filter.

Example 3.1 *Consider the following two-dimensional HP-filter: Let the same shift matrix P given in (2.4) act on each of the trends $y^i = (y_1^i, \ldots, y_T^i),\ i = 1, 2$.*
We have,

$$x^i = y^i + u^i,\ Py^i = v^i,\ \ i = 1, 2,$$

where, the noises u^1 and u^2, respectively v^1 and v^2, may be correlated.
With, $x = (x^1, x^2)$, $y = (y^1, y^2)$, $u = (u^1, u^2)$, $v = (v^1, v^2)$ and

$$A := \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix},$$

we get the following two-dimensional HP-filter:

$$x = y + u,\ \ Ay = v. \tag{3.2}$$

We suggest the following multivariate HP-filter.

Definition 3.2 *A multivariate HP-filter decomposes* $x \in \mathbb{R}^{dT}$ *into a sum of a nonstationary trend* $y \in \mathbb{R}^{dT}$ *and a cyclical residual component (noise term)* $u \in \mathbb{R}^{dT}$ *by minimizing the following weighted sum*

$$(x-y)'\Sigma^{-1}(x-y) + y'A'\Omega^{-1}Ay, \tag{3.3}$$

with respect to y*, where* (Σ, Ω) *is a pair of positive definite matrices with appropriate dimensions that acts as a smoothing parameter.*

Similarly to the shift operator P, we assume that for some fixed $0 < k < T$, the trend-smoothing operator A is a $d(T-k) \times dT$-matrix, with rank $d(T-k)$, making the matrix AA' invertible.

Setting $d = 1, k = 2$, $A = P$, $\Sigma = I_T$ and $\Omega^{-1} = \alpha I_{T-2}$, (3.3) reduces to (2.2). When $\Omega^{-1} = \mathrm{diag}\{\alpha_1, \ldots, \alpha_{T-2}\}$, we get the model suggested in Razzak and Dennis (1995) (see also Reeves *et al.* (2000)).

The signal $v := Ay$ does not determine a unique y but rather the set of solutions

$$y := \{A'(AA')^{-1}v + Z\gamma; \;\; \gamma \in \mathbb{R}^{dk}\} \tag{3.4}$$

where, the $dT \times dk$-matrix Z satisfies

$$\Pi + ZZ' = I_{dk} \qquad AZ = 0, \qquad Z'Z = I_{dk}, \tag{3.5}$$

with I_{dk} denoting the $dk \times dk$ identity matrix and

$$\Pi = A'(AA')^{-1}A \tag{3.6}$$

is an orthogonal projector associated with A i.e. it satisfies $\Pi^2 = \Pi$.

In view of Eq. (3.1), the time series x can be represented in terms of (u, v) as

$$x = u + A'(AA')^{-1}v + Z\gamma, \tag{3.7}$$

for some $\gamma \in \mathbb{R}^{dk}$.

Since the matrix $(I_{dT} + \Sigma A'\Omega^{-1}A)$ is invertible, the unique solution $\hat{y}$ to the optimal problem (3.3) is

$$\hat{y} := y\left((\Sigma, \Omega), x\right) = \left(I_{dT} + \Sigma A'\Omega^{-1}A\right)^{-1} x, \tag{3.8}$$

where I_{dT} denotes the $dT \times dT$ identity matrix. Eq. (3.8) defines the descriptive filter that associates a trend y to the time series x, depending on the smoothing parameter (Σ, Ω) and the disturbance operator A.

As for the univariate HP filter, assume the noise term u and the signal term v have zero mean and general covariance matrices Σ_u and Σ_v, respectively:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}(0, \Sigma_{uv}). \tag{3.9}$$

with covariance matrix

$$\Sigma_{uv} := \begin{pmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{pmatrix},$$

which, by (3.4), makes the trend y following a Gaussian process. This turns the time series x into a trend y generated by a Gaussian process and a Gaussian disturbance u. That is, in view of (3.7), (x, y) is normally distributed:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} Z \\ Z \end{pmatrix} \gamma, \Sigma_{xy} \right), \tag{3.10}$$

with covariance matrix

$$\Sigma_{xy} := \begin{pmatrix} \Sigma_u + Q_v & Q_v \\ Q_v & Q_v \end{pmatrix},$$

where,

$$Q_v := A'(AA')^{-1}\Sigma_v(AA')^{-1}A \tag{3.11}$$

is a symmetric matrix that, which in view of (3.5), satisfies $Q_v Z = 0$.

Now, thanks to (3.10), an explicit expression of the conditional expectation of the trend y given the time series x reads:

$$E[\,y|\,x] = Z\gamma + Q_v \left[\Sigma_u + Q_v\right]^{-1} (x - Z\gamma). \tag{3.12}$$

In a similar manner as in Section 3.1, we suggest to find a multivariate version of the univariate noise-to-signal 'ratio' i.e. the best smoothing parameters (Σ, Ω) that minimize the quadratic gap between $E[\,y|\,x]$ and $y\,((\Sigma, \Omega), x)$. We will proceed in two steps. First, we apply Schlichts criterium to find the best smoothing parameters and then show that this criterium is indeed optimal i.e. these matrices do minimize the quadratic gap between $E[\,y|\,x]$ and $y\,((\Sigma, \Omega), x)$.

3.1 A Criterium for choosing the best smoothing parameter

Adapting Criterium (2.17) to the multivariate case, a best smoothing parameter for the multivariate HP filter is a pair of positive definite matrices (Σ, Ω) for which

$$y\,((\Sigma, \Omega), x) - Z\hat{\gamma} = E[y\,|\,x] - Z\gamma, \text{ for all realisations } x \text{ in } \mathbb{R}^{dT}, \tag{3.13}$$

which is equivalent to

$$\Sigma A'\Omega^{-1}\Sigma_v(AA')^{-1}A = \Sigma_u - Z\left(Z'[\Sigma_u + Q_v]^{-1}Z\right)^{-1} Z', \tag{3.14}$$

where, $\hat{\gamma}$ is the maximum likelihood estimator of γ based on observations from the time series x:

$$\hat{\gamma} = \arg\min_{\gamma} (x - Z\gamma)' \left[\Sigma_u + Q_v\right]^{-1} (x - Z\gamma),$$

which is explicitly given by

$$\hat{\gamma} = (Z'[\Sigma_u + Q_v]^{-1}Z)^{-1}Z'[\Sigma_u + Q_v]^{-1}x; \tag{3.15}$$

the matrix $Z'[\Sigma_u + Q_v]^{-1}Z$ being positive definite.

We note that if (Σ, Ω) are solutions of (3.14), then by mutiplying both sides of the equation with Z, we necessarily get

$$\left(Z' \left[\Sigma_u + Q_v\right]^{-1} Z\right)^{-1} = Z'\Sigma_u Z.$$

In this case (3.14) reduces to

$$\Sigma A'\Omega^{-1}\Sigma_v(AA')^{-1}A = \Sigma_u\Pi + \Pi\Sigma_u(I_{dT} - \Pi), \tag{3.16}$$

where, as above, $\Pi = A'(AA')^{-1}A$.

In Proposition (3.4) below, we will show that the solvability of Eq. (3.14) is equivalent to imposing $\Pi\Sigma_u(I_{dT} - \Pi) = 0$ which means the vectors Πu and $(I_{dT} - \Pi)u$ are uncorrelated.

In the next lemma we derive some consequences of imposing this condition that turns out very useful for the sequel.

Lemma 3.3 *We have*

(1)
$$\Pi\Sigma_u(I_{dT} - \Pi) = 0 \quad \textit{if and only if} \quad \Pi\Sigma_u = \Sigma_u\Pi. \tag{3.17}$$

(2) *Under (3.17), we have*

$$\left(Z'\Sigma_u Z\right)^{-1} = Z' \left(\Sigma_u\right)^{-1} Z, \tag{3.18}$$

$$Z' \left[\Sigma_u + Q_v\right]^{-1} Z = Z' \left(\Sigma_u\right)^{-1} Z, \tag{3.19}$$

$$\left(Z' \left[\Sigma_u + Q_v\right]^{-1} Z\right)^{-1} = Z'\Sigma_u Z. \tag{3.20}$$

and

$$Z\hat{\gamma} = ZZ'x. \tag{3.21}$$

(3) $\Pi\Sigma_u = \Sigma_u\Pi$ *holds if and only if*

$$\left(I_{dT} + \Sigma_u A'\Sigma_v^{-1}A\right)^{-1} = ZZ' + Q_v \left[\Sigma_u + Q_v\right]^{-1}. \tag{3.22}$$

Proof. The fact that $\Pi\Sigma_u = \Sigma_u\Pi$ implies $\Pi\Sigma_u(I_{dT} - \Pi) = 0$, is immediate. Assume $\Pi\Sigma_u(I_{dT} - \Pi) = 0$. Then $\Pi\Sigma_u Z = 0$, which in turn yields that $\Sigma_u Z \in Ker(\Pi)$. Since Σ_u is invertible, we get

$$\Sigma_u(Ker(\Pi)) = Ker(\Pi). \tag{3.23}$$

On the other hand, we also have $(I_{dT} - \Pi)\Sigma_u \Pi = 0$. This gives $(I_{dT} - \Pi)\Sigma_u a = 0$ for all $a \in Im(\Pi)$. Hence,

$$\Sigma_u(Im(\Pi)) = Im(\Pi). \tag{3.24}$$

Now, (3.23) and (3.24) yield $\Pi\Sigma_u = \Sigma_u\Pi$.

We note that (3.20) follows from (3.18) and (3.19). Relation (3.18) follows from $\Pi\Sigma_u = \Sigma_u\Pi$. To show (3.19), we apply the following formula (see Rao (1965), Exercise 2.9, pp. 33)

$$(C + EDE')^{-1} = C^{-1} - C^{-1}E(E'C^{-1}E + D^{-1})^{-1}E'C^{-1}$$

to $[\Sigma_u + Q_v]^{-1}$. We get

$$[\Sigma_u + Q_v]^{-1} = \Sigma_u^{-1} - \Sigma_u^{-1}A'(AA')^{-1}\left[(AA')^{-1}A\Sigma_u^{-1}A'(AA')^{-1} + \Sigma_v^{-1}\right]^{-1}(AA')^{-1}A\Sigma_u^{-1}.$$

Now, since by (3.23), $\Sigma_u(Ker(\Pi)) = Ker(\Pi)$, multiplying both sides with Z' from the left and with Z from the right, we get (3.19).

Relations (3.21) and (3.22) are straightforward. □

The following proposition characterizes the solvability of (3.14) i.e. the set of multivariate 'noise-to-signal ratio's'.

Proposition 3.4 *Equation (3.14) is solvable if and only if*

$$\Pi\Sigma_u(I_{dT} - \Pi) = 0. \tag{3.25}$$

In this case, Σ, Ω satisfy (3.14) if and only if

$$\Sigma A'\Omega^{-1}A = \Sigma_u A'\Sigma_v^{-1}A. \tag{3.26}$$

In particular,

$$y\left((\Sigma, \Omega), x\right) = y\left((\Sigma_u, \Sigma_v), x\right), \tag{3.27}$$

for all $x \in \mathbb{R}^{dT}$. Hence, this family of matrices (Σ, Ω) is an equivalence class, that gives the same value to the trend.

Moreover,

$$\hat{y}(\Sigma, \Omega) - E(y \mid x) = ZZ'(x - Z\gamma),$$

and its covariance matrix is

$$cov\left(E[\,y|\,x] - y\left((\Sigma, \Omega), x\right)\right) = \Sigma_u ZZ'. \tag{3.28}$$

In particular,

$$E\left(\|E[\,y|\,x] - y\left((\Sigma, \Omega), x\right)\|^2\right) = E\left(\|ZZ'(x - Z\gamma)\|^2\right) = trace(\Sigma_u ZZ'). \tag{3.29}$$

Proof. Assume (Σ, Ω) solves (3.14). Then, we get Eq. (3.16):

$$\Sigma A'\Omega^{-1}\Sigma_v(AA')^{-1}A - \Sigma_u\Pi = \Pi\Sigma_u(I_{dT} - \Pi).$$

Multiplying both sides with $(I_{dT} - \Pi)c$, for an arbitrary $c \in \mathbb{R}^{dT}$, and using $A\Pi = A$, we get

$$\Pi\Sigma_u(I_{dT} - \Pi)c = \left(\Sigma A'\Omega^{-1}\Sigma_v(AA')^{-1}A - \Sigma_u\Pi\right)(I_{dT} - \Pi)c = 0,$$

arriving at $\Pi\Sigma_u(I_{dT} - \Pi) = 0$.

Conversely, assuming (3.25), Eq. (3.14) (or (3.16)) reduces to

$$\Sigma A'\Omega^{-1}\Sigma_v(AA')^{-1}A = \Sigma_u\Pi,$$

for which (Σ_u, Σ_v) and $(\Sigma_u + ZZ', \Sigma_v)$ are solutions.

The rest of the proof is straightforward. □

Remark 3.5

(*a*) Since $Z'A' = 0$, if (Σ, Ω) solves (3.26), then $(\Sigma ZZ', \Omega)$ is also a solution.

(*b*) Using the fact that $\Pi A' = A'$, in view of (3.27) we also get that

$$y\left((\Sigma, \Omega), x\right) = y\left((\Pi\Sigma_u\Pi, \Sigma_v), x\right) = y\left((\Sigma_u, \Sigma_v), x\right), \tag{3.30}$$

for all $x \in \mathbb{R}^{dT}$, which involves only the variance of Πu which is $\Pi\Sigma_u\Pi$ and Σ_v, instead of the whole covarinace matrix Σ_u.

We end this section with several examples of positive definite matrices that satisfy (3.26) i.e. being a 'noise-to-signal ratio'.

Example 3.6

(*a*) (Σ_u, Σ_v), $(\Sigma_u + ZZ', \Sigma_v)$ *and* $(\Pi\Sigma_u\Pi + (I - \Pi)L(I - \Pi), \Sigma_v)$, *where,* $\Pi\Sigma_u\Pi$ *is the covariance matrix of* Πu, *and* L *is any positive matrix.*

(*b*) *The pair of matrices* (Σ, Σ_v) *satisfies (3.26), for any positive definite matrix* Σ *that satisfies*

$$\Sigma\Pi = \Pi\Sigma,$$

which is equivalent to choosing Σ *of the form*

$$\Sigma = \begin{bmatrix} var(\Pi u) & 0 \\ 0 & L \end{bmatrix}, \tag{3.31}$$

written in the basis

$$(b_i) := A'g_1/\sqrt{\lambda_1}, \ldots, A'g_{d(T-k)}/\sqrt{\lambda_{d(T-k)}}, Z_1, \ldots, Z_{kd}, \tag{3.32}$$

where λ_i *are the eigenvalues of* AA' *and* $\{g_1, \ldots, g_{d(T-k)}\}$ *an orthonormal set of eigenvectors corresponding to* λ_i, $i = 1, \ldots, \lambda_{d(T-k)}$. *The matrix* L *is any arbitrary* $dk \times dk$*-matrix. When* $L = var((I - \Pi)u)$, $\Sigma = \Sigma_u$.

(c) *In the particular case where* $\Omega = \Sigma_v = \sigma_v^2 I_{d(T-k)}$, *we have*

$$\hat{y} = \left(I_{dT} + \frac{\Sigma_u}{\sigma_v^2} A'A \right)^{-1} x, \tag{3.33}$$

suggesting the noise-to-signal ratio $\alpha = \Sigma_u / \sigma_v^2$, *which is a matrix.*

Let

$$\hat{y} = \sum_{i=1}^{d(T-k)} \hat{y}_i \frac{A' g_i}{\sqrt{\lambda_i}} + \sum_{i=d(T-k)+1}^{kd} \hat{y}_i Z_{i-d(T-k)}$$

be the decomposition of $\hat{y}$ *in the basis* $(b_i)_i$ *given by (3.32), where,* $\hat{y}_i =< \hat{y}, b_i >$.

From (3.33) we derive

$$\sum_{i=1}^{d(T-k)} (\hat{y}_i + \hat{y}_i \lambda_i \frac{\Sigma_u}{\sigma_v^2}) \frac{A' g_i}{\sqrt{\lambda_i}} + \sum_{i=d(T-k)+1}^{dT} \hat{y}_i Z_{i-d(T-k)} = x.$$

In particular, if $\Sigma_u A' g_i = \sigma_i^2 A' g_i$ *for all* $i = 1, \ldots, d(T-k)$, *then*

$$x_i = \hat{y}_i, \quad \textit{for all } i > d(T-k)$$

and

$$x_i = \hat{y}_i (1 + \lambda_i \frac{\sigma_i^2}{\sigma_v^2}), \quad i = 1, \ldots, d(T-k) \tag{3.34}$$

where, $x_i =< x, b_i >$.

(d) *The multivariate HP filter suggested in Razzak and Dennis (1995) corresponds to the case where* $d = 1, k = 2$, $A = P$, $\Sigma_u = \sigma_u^2 I_T$, $\Sigma = I_T$, $\Sigma_v = diag\{\sigma_v^2(1), \ldots, \sigma_v^2(T-2)\}$ *and* $\Omega^{-1} = diag\{\alpha_1, \ldots, \alpha_{T-2}\}$. *Plugging in these matrices in Eq. (3.26), we get that*

$$\Omega^{-1} = diag\{\alpha_1, \ldots, \alpha_{T-2}\} = \sigma_u^2 \Sigma_v^{-1},$$

or, equivalently,

$$\alpha_t = \sigma_u^2 / \sigma_v^2(t), \quad t = 1, \cdots, T-2.$$

3.2 Optimality of the best smoothing parameters

In this section we will show that the matrices (Σ, Ω) that satisfy Eq. (3.26) (or Eq. (3.14)) minimize the gap between $y((\Sigma, \Omega), x)$ and $E[y|x]$, over an appropriate class $\mathcal{A}$ of admissible matrices in the sense that

$$\arg \min_{(\Sigma,\Omega)\in\mathcal{A}} ||\hat{y}(\Sigma, \Omega) - E(y \,|\, x)|| = \{(\Sigma_*, \Omega_*) \in \mathcal{A} : \Sigma_* A' \Omega_*^{-1} A = \Sigma_u A' \Sigma_v^{-1} A\}.$$

By Proposition 3.4, in this class $\mathcal{A}$ of such admissible matrices we expect that

$$\|y\left((\Sigma,\Omega),x\right)-E[\,y|\,x]\| \geq \|y\left((\Sigma_*,\Omega_*),x\right)-E[\,y|\,x]\| = \|ZZ'(x-Z\gamma)\|. \quad (3.35)$$

We have, with $\hat{y} := y\left((\Sigma,\Omega),x\right)$,

$$\hat{y}-E[\,y|\,x] = Z(\hat{\gamma}-\gamma)+\hat{y}-Z\hat{\gamma}-Q_v[\Sigma_u+Q_v]^{-1}(x-Z\gamma),$$

where, by Lemma 3.3, $Z\hat{\gamma}=ZZ'x$.

Let $<,>$ denote the scalar product in $\mathbb{R}^{dT}$ and set

$$\Phi(\Sigma,\Omega,x,\gamma) = \|\hat{y}-Z\hat{\gamma}-Q_v[\Sigma_u+Q_v]^{-1}(x-Z\gamma)\|^2+2<Z(\hat{\gamma}-\gamma),\hat{y}-Z\hat{\gamma}>,$$

which is of the form

$$\Phi(\Sigma,\Omega,x,\gamma) =< B_1x,x> + < B_2x,\Gamma>,$$

where,

$$B_2 = ZZ'(I_{dT}+\Sigma A'\Omega^{-1}A)^{-1}-ZZ'$$

and B_1 and Γ can be given explicitly, but omit them for simplicity.

Obviously, (3.35) holds if and only if

$$\Phi(\Sigma,\Omega,x,\gamma)\geq 0$$

for all $x\in\mathbb{R}^{dT}$. But, this is true if and only if $B_1\geq 0$ and $B_2=0$.

Now, $B_2=0$ is equivalent to $ZZ'\Sigma A'\Omega^{-1}A=0$, which holds if and only if $ZZ'\Sigma = \Sigma ZZ'$, due to the fact that $Im(A'\Omega^{-1}A)=Im(\Pi)$, or, using (3.5), if and only if

$$\Sigma\Pi = \Pi\Sigma. \quad (3.36)$$

But, it is easy to check that (3.36) holds if and only if

$$\hat{y} = ZZ'x+(I_{dT}-ZZ')(I_{dT}+\Sigma A'\Omega^{-1}A)^{-1}x. \quad (3.37)$$

This yields that

$$\hat{y}-E[\,y|\,x] = ZZ'(x-Z\gamma)+(I_{dT}-ZZ')\left((I_{dT}+\Sigma A'\Omega^{-1}A)^{-1}x-Q_v[\Sigma_u+Q_v]^{-1}(x-Z\gamma)\right).$$

Thus, since ZZ' and $I_{dT}-ZZ'$ are orthogonal, we get

$$\|\hat{y}-E[\,y|\,x]\|^2 = \|ZZ'(x-Z\gamma)\|^2+\|(I_{dT}-ZZ')\left((I_{dT}+\Sigma A'\Omega^{-1}A)^{-1}x-Q_v[\Sigma_u+Q_v]^{-1}(x-Z\gamma)\right)\|^2,$$

arriving at (3.35).

Now, if

$$\Sigma\Pi \neq \Pi\Sigma, \quad (3.38)$$

then, there exists $x\in\mathbb{R}^{dT}$ such that $\Phi(\Sigma,\Omega,x,\gamma)<0$. In this case

$$\|y\left((\Sigma,\Omega),x\right)-E[\,y|\,x]\| < \|ZZ'(x-Z\gamma)\|. \quad (3.39)$$

We have proved the following

Proposition 3.7

(*a*) $\Phi(\Sigma, \Omega, x, \gamma) \geq 0$ *for all* x *if and only if* $\Sigma\Pi = \Pi\Sigma$.
In this case

$$\|y\left((\Sigma, \Omega), x\right) - E[\, y|\, x]\| \geq \|ZZ'(x - Z\gamma)\|. \tag{3.40}$$

(*b*) *If* $\Sigma\Pi \neq \Pi\Sigma$, *there exists* $x \in \mathbb{R}^{dT}$ *such that* $\Phi(\Sigma, \Omega, x, \gamma) < 0$, *in which case*

$$\|y\left((\Sigma, \Omega), x\right) - E[\, y|\, x]\| < \|ZZ'(x - Z\gamma)\|. \tag{3.41}$$

To sum up, an appropriate class $\mathcal{A}$ of admissible matrices is

$$\mathcal{A} = \{(\Sigma, \Omega),\ \text{positive definite s.t. } \Sigma\Pi = \Pi\Sigma\}$$

and the optimal smoothing parameters are those matrices $(\Sigma, \Omega) \in \mathcal{A}$ that solve Eq. (3.26).

We have the following multivariate version of Proposition (2.19), that constitutes the main result of the paper.

Theorem 3.8 *Assume* $\Sigma_u\Pi = \Pi\Sigma_u$. *Then,*

$$\arg \min_{(\Sigma,\Omega)\in\mathcal{A}} ||\hat{y}(\Sigma, \Omega) - E(y\,|\,x)|| = \{(\Sigma, \Omega) \in \mathcal{A} : \Sigma A'\Omega^{-1}A = \Sigma_u A'\Sigma_v^{-1}A\}.$$

Moreover, for $(\Sigma, \Omega) \in \mathcal{A}$, *the optimal gap is*

$$\hat{y}(\Sigma, \Omega) - E(y\,|\,x) = ZZ'(x - Z\gamma),$$

and its covariance matrix is

$$cov\left(E[\, y|\, x] - y\left((\Sigma, \Omega), x\right)\right) = \Sigma_u ZZ'. \tag{3.42}$$

In particular,

$$E\left(\|E[\, y|\, x] - y\left((\Sigma, \Omega), x\right)\|^2\right) = E\left(\|ZZ'(x - Z\gamma)\|^2\right) = trace(\Sigma_u ZZ'). \tag{3.43}$$

References

[1] Apel, M., Hansen, J. and Lindberg, H. (1996): Potential output and output gap. *Quarterly Review III- Sveriges Riksbank.*

[2] Anselone P. M., Laurent P. J. (1968): A general method for construction of interpolating or smoothing spline-functions, *Numer. Math.* 12 (1), 66-82.

[3] Araujo, F. Areaso, M. B. M. and Neto, J. A. R. (2003): r-filters: a Hodrick-Prescott Filter generalization. *Working Paper Series (69), Banco Central Do Brasil.*

[4] Dermoune, A., Djehiche, B. and Rahmania, N. (2007): Consistent estimators of the smoothing parameter in the Hodrick-Prescott Filter. *J. Japan Statist. Soc.* (to appear).

[5] Harvey, A. C. (1986): Trend and cycles in macroeconomic time series. *Journal of Business and Economic Statistics*, Vol. 3 (3), pp. 216-227.

[6] Hodrick, R. and Prescott, E. C. (1980): Postwar U.S. business cycles: An empirical investigation. *Discussion paper* 451, Carnegie-Mellon University.

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

[8] Kydland, E. F., Prescott, E. C. (1990): Business cycles: Real facts and monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 14 (Spring), 3-18.

[9] C. R. Rao (1965): *Linear Statistical Inference and its Applications*, Wiley, New York.

[10] Razzak, W. and Dennis, R. J. (1995): Estimates of New Zealand's output gap using the Hodrick-Prescott filter with a non-constant smoothing parameter. *Discussion paper G95/8, Reserve Bank of New Zealand.*

[11] Razzak, W. (1997): The Hodrick-Prescott technique: A smoother versus a filter. An application to New Zealand GDP. *Economic Letters*, (57), 163-168.

[12] Reeves, J. J., Blyth, C. M., Triggs, C. M. and Small, J. P. (2000): The Hodrick-Prescott Filter, a generalization and a new procedure for extracting an empirical cycle from a series. *Studies in Nonlinear Dynamics and Econometrics.* 4(1): 1-16.

[13] Schlicht, E. (2006): Estimating the smoothing parameter in the so-called Hodrick-Prescott filter. *J. Japan Statist. Soc.*, Vol. 35 No. 1, 99-119.

[14] Schlicht, E. (1984): Seasonal adjustment in a stochastic model. *Statistische Hefte*, 25, 1-12.

[15] Trimbur, T.M., (2006): Detrending economic time series: A Bayesian generalization of the Hodrick-Prescott Filter, J. Forecast. 25: 247-273.