On the functional Hodrick-Prescott filter with compact operators and non-compact operators

Hiba Nassar

Joint work with Boualem Djehiche and Astrid Hilbert.

LinStat 2014
August 24 - 28, 2014

Outline

The univariate HP filter extracts a 'signal' $y(\alpha, x) = (y_1(\alpha, x), \ldots, y_T(\alpha, x))$ from a noisy time series $x = (x_1, \ldots, x_T)$ as a minimizer of

$$\sum_{t=1}^{T}(x_t - y_t)^2 + \alpha \sum_{t=3}^{T}(y_t - 2y_{t-1} + y_{t-2})^2, \qquad (1)$$

with respect to $y = (y_1, \ldots, y_T)$, for an appropriately chosen positive parameter α, called the smoothing parameter.
The second order diferencing operator $Py(t) = y_t - 2y_{t-1} + y_{t-2}$ is written in vector form as the following $(T-2) \times T$-matrix

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

To determine an appropriate value of the smoothing parameter α, Hodrick and Prescott (1997) suggest the time series (x, y) satisfies the following linear mixed model:

$$\begin{cases} x = y + u, \\ Py = v. \end{cases} \tag{2}$$

where, $u \sim N(0, \sigma_u^2 I_T)$ and $v \sim N(0, \sigma_v^2 I_{T-2})$.
The 'optimal smooth' signal associated with x is

$$\bar{y}(\alpha, x) := \arg\min_y \left\{ \|x - y\|^2_{\mathbb{R}^T} + \alpha \|Py\|^2_{\mathbb{R}^{T-2}} \right\}. \tag{3}$$

Using the model above, the appropriate smoothing parameter turns out to be the noise-to-signal ratio $\alpha^* = \sigma_u^2 / \sigma_v^2$.

Schlicht in (2005) proved that the noise-to-signal ratio satisfies

$$E[y|x] = y\left(\frac{\sigma_u^2}{\sigma_v^2}, x\right), \tag{4}$$

where $E[y|x]$ is the best predictor of any signal y given the time series x. Dermoune *et al.* proposed in (2009) an optimality criterion for choosing the smoothing parameter for the HP-filter. The smoothing parameter α is chosen as the following:

$$\alpha^* = \arg\min_{\alpha} \left\{ \|E[y|x] - y(\alpha, x)\|^2 \right\} \tag{5}$$

Furthermore, Dermoune *et al.* (2009) proposed a multivariate version of the HP filter and determined the possible optimal smoothing parameters.

Definition

Let H_1 and H_2 be two separable Hilbert spaces, with norms $\|\cdot\|_{H_i}$ and inner products $\langle\cdot,\cdot\rangle_{H_i}$, $i = 1, 2$, and $x \in H_1$ be a functional time series of observables. A functional Hodrick-Prescott filter reconstructs an 'optimal smooth signal' $y \in H_1$ that solves an equation $Ay = v$, corrupted by a noise v which is apriori unobservable, from observations x corrupted by a noise u which is also apriori unobservable:

$$\begin{cases} x = y + u, \\ Ay = v, \end{cases} \tag{6}$$

given the linear operator $A : H_1 \longrightarrow H_2$ and u, v are independent random variables with zero mean and covariance operators Σ_u and Σ_v respectively.

The 'optimal smooth' signal associated with x is given by:

$$y(B,x) := \arg\min_y \left\{ \|x-y\|_{H_1}^2 + \langle Ay, BAy\rangle_{H_2} \right\}, \tag{7}$$

where $B : H_2 \longrightarrow H_2$ is a smoothing operator, provided that

$$\langle Ah, BAh\rangle_{H_2} \geq 0, \quad h \in H_1.$$

Definition

The optimal smoothing operator associated with the Hodrick-Prescott filter (6) is the minimizer of the difference between the optimal solution $y(B,x)$, and the conditional expectation $E[y|x]$, the best predictor of any signal y given the functional data x:

$$\hat{B} = \arg\min_B \|E[y|x] - y(B,x)\|_{H_1}^2 . \tag{8}$$

Proposition

Let $A : H_1 \longrightarrow H_2$ *be a compact operator with the singular system* (λ_n, e_n, d_n)*. Assume further that the smoothing operator* $B : H_2 \longrightarrow H_2$ *is linear, bounded and satisfies*

$$\langle Ah, BAh \rangle_{H_2} \geq 0, \quad h \in H_1. \tag{9}$$

Then, there exists a unique $y(B, x) \in H_1$ *which minimizes the functional*

$$J_B(y) = \|x - y\|^2_{H_1} + \langle Ay, BAy \rangle_{H_2}.$$

This minimizer is given by the formula

$$y(B, x) = (I_{H_1} + A^* BA)^{-1} x. \tag{10}$$

If the smoothing operator $B : H_2 \to H_2$ admits the following representation

$$Bh = \sum_{k=1}^{\infty} \beta_k \langle h, d_k \rangle d_k, \quad h \in H_2, \tag{11}$$

where $\beta_k > 0, \; k = 1, 2, \ldots$, and the sum converges in the operator norm, i.e. B is linear, compact and injective, then

$$y(B, x) = (I_{H_1} + A^* BA)^{-1} x = \sum_{j=1}^{\infty} \frac{1}{1 + \lambda_j^2 \beta_j} \langle x, e_j \rangle e_j. \tag{12}$$

Assumptions

1. u and v are independent random variables with zero mean and covariance operators Σ_u and Σ_v respectively.
2. The independent random variables u and v are respectively $N(0, \Sigma_u)$ and $N(0, \Sigma_v)$ distributed, where the covariance operators Σ_u and Σ_v are positive-definite and trace class operators on H_1 and H_2 respectively.
3. The orthogonal (in H_1) random variables Πu and $(I_{H_1} - \Pi)u$ are independent:
$$\Pi\Sigma_u = \Sigma_u\Pi. \quad (13)$$
4. The operator
$$Q_v := A^*(AA^*)^{-1}\Sigma_v(AA^*)^{-1}A$$
is trace class.

Since the covariance operators Σ_u and Σ_v are trace class and thus compact, by Riesz' Representation Theorem, they admit the following decompositions:

$$\Sigma_u h = \sum_{k=1}^{\infty} \mu_k \langle h, e_k \rangle e_k, \qquad h \in H_1, \tag{14}$$

$$\Sigma_v h = \sum_{k=1}^{\infty} \tau_k \langle h, d_k \rangle d_k, \qquad h \in H_2, \tag{15}$$

where the sums converge in the corresponding operator norm.

Proposition

Let X, Y be jointly Gaussian H-valued random variables. Assume that both X and Y have means μ_X and μ_Y, and that the covariance of X, Σ_X, is injective. Then, the conditional expectation of Y given X is

$$E[Y|X] = \mu_Y + \Sigma_{XY}\Sigma_X^{-1}(X - \mu_X), \tag{16}$$

provided that the operator

$$T = \Sigma_{XY}\Sigma_X^{-\frac{1}{2}} \tag{17}$$

is Hilbert-Schmidt.

See Mandelbaum [8]

Theorem

Let Assumptions (1) to (4) hold, and that

$$\|T\|_2^2 = \sum_{k=1}^{\infty} \frac{\tau_k}{\lambda_k^2} \left(\frac{\lambda_k^2 \mu_k}{\tau_k} + 1 \right)^{-1} < \infty, \tag{18}$$

then, for all $x \in H_1$, the smoothing operator (which is linear, compact and injective)

$$\hat{B}h := (AA^*)^{-1} A \Sigma_u A^* \Sigma_v^{-1} h = \sum_{k=1}^{\infty} \frac{\mu_k}{\tau_k} \langle h, d_k \rangle d_k, \qquad h \in H_2, \tag{19}$$

where, the sum converges in the operator norm, is the unique operator which satisfies

$$\hat{B} = \arg\min_{B} \|y(B,x) - E[y|x]\|_{H_1},$$

where the minimum is taken with respect to all linear bounded operators which satisfy the positivity condition (9).

Furthermore, we have

$$y(\hat{B},x) - E[y|x] = (I_{H_1} - \Pi)(x - E[x]), \tag{20}$$

and its covariance operator is

$$\mathrm{cov}\,(y(\hat{B},x) - E[y|x]) = (I_{H_1} - \Pi)\Sigma_u. \tag{21}$$

In particular,

$$E\left(\left\|y(\hat{B},x) - E[y|x]\right\|_{H_1}^2\right) = \mathrm{trace}\left((I_{H_1} - \Pi)\Sigma_u\right). \tag{22}$$

Assume that $u \sim N(0, \Sigma_u)$ and $v \sim N(0, \Sigma_v)$ where Σ_u and Σ_v are self-adjoint positive-definite and bounded but not trace class operators on H_1 and H_2, respectively.

Following Rozanov (1968), we can look at these Gaussian variables as generalized random variables on an appropriate Hilbert scale, where the covariance operators can be maximally extended to self-adjoint positive-definite, bounded and trace class operators on a larger space.

We first construct the Hilbert scales $(H_1^n)_{n\in\mathbb{R}}$ $((H_2^n)_{n\in\mathbb{R}})$ induced by $K_1 = (A^*A)^{-1}$ $(K_2 = (AA^*)^{-1}\ )$ of H_1, (H_2).

For all $n \in \mathbb{N}$ the space H_1^n is a complete space with respect to the norm induced by the following inner product

$$\langle x, y\rangle_{H_1^n} := \langle (A^*A)^{-n}x, (A^*A)^{-n}y\rangle_{H_1}, \quad x, y \in H_1^n. \tag{23}$$

Assumption (5): There is $n_0 > 0$ such that for all $n \geq n_0$ we have

$$\sum_{k=1}^{\infty} \lambda_k^{4n-2} \mu_k < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k^{4n} \tau_k < \infty.$$

Under Assumption 5, the covariance operators $\tilde{\Sigma}_u, \tilde{\Sigma}$ and $\tilde{\Sigma}_v$ are trace class on the Hilbert spaces H_1^{-n} and H_2^{-n}, respectively, where

$$\tilde{\Sigma}_u = (A^*A)^n \Sigma_u (A^*A)^n, \quad \tilde{\Sigma}_v = (AA^*)^n \Sigma_v (AA^*)^n \tag{24}$$

and

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_u + \widetilde{Q}_v & \widetilde{Q}_v \\ \widetilde{Q}_v & \widetilde{Q}_v \end{pmatrix}, \tag{25}$$

where,

$$\widetilde{Q}_v := (A^*A)^n Q_v (A^*A)^n = A^*(AA^*)^{-1} \widetilde{\Sigma}_v (AA^*)^{-1} A. \tag{26}$$

Theorem

Let Assumption 5 hold. Then, the operator

$$\hat{B}h := (AA^*)^{-1}A\tilde{\Sigma}_u A^*\tilde{\Sigma}_v^{-1}h, \quad h \in H_2^{-n}, \tag{27}$$

is the unique optimal smoothing operator associated with the HP filter associated with H_1^{-n}*-valued data* x*.*

In this section we apply Theorem 7 to the case where u and v are white noise i.e. u and v are independent Gaussian random variables with zero means and covariance operators

$$\Sigma_u = \sigma_u I_{H_1}, \qquad \Sigma_v = \sigma_v I_{H_2},$$

where I_{H_1} and I_{H_2} denotes the H_1 and H_2 identity operators, respectively and σ_u and σ_v are constant scalars. Assumption 5, reduces to

Assumption 6. There is an $n_0 > 0$ such that $\sum_{k=1}^{\infty} \lambda_k^{2(2n-1)} < \infty$ for all $n \geq n_0$.

Under this assumption, the associated covariance operators $\tilde{\Sigma}_u, \tilde{\Sigma}_v$ and $\widetilde{Q}_v$ are all trace class operators. Hence, the expression (27) giving the optimal smoothing operator $\hat{B}$ reduces to

$$\hat{B} = (AA^*)^{-1} A \Sigma_u A^* \Sigma_v^{-1} = \frac{\sigma_u}{\sigma_v} \sum_{k=1}^{\infty} \langle \cdot, d_k \rangle d_k = \frac{\sigma_u}{\sigma_v} I_{H_2^{-n}}, \tag{28}$$

i.e. $\hat{B}$ is the noise-to-signal ratio.

Assumptions

7. the linear operator $A : H_1 \to H_2$ is
   - (1a) Closed and defined on a dense subspace $\mathcal{D}(A)$ of H_1,
   - (1b) Its range, $\mathrm{Ran}(A)$, is closed.
8. u and v are independent Gaussian random variables with zero mean and covariance operators Σ_u and Σ_v respectively.
9. The orthogonal (in H_1) random variables Πu and $(I_{H_1} - \Pi)u$ are independent:

$$\Pi\Sigma_u = \Sigma_u\Pi. \tag{29}$$

Assumption (1) is equivalent to the fact that $A^\dagger$ is bounded.

Proposition

Let $A : H_1 \longrightarrow H_2$ be a closed, linear operator and its domain is dense in H_1. Assume further the smoothing operator $B : H_2 \longrightarrow H_2$ is closed, densely defined and satisfies

$$\langle Ah, BAh \rangle_{H_2} \geq 0, \quad h \in H_1. \tag{30}$$

Then, there exists a unique $y(B,x) \in H_1$ which minimizes the functional

$$J_B(y) = \|x - y\|_{H_1}^2 + \langle Ay, BAy \rangle_{H_2}.$$

This minimizer is given by the formula

$$y(B,x) = (I_{H_1} + A^* BA)^{-1} x. \tag{31}$$

Given Assumption (8), it holds that (x, y) is Gaussian with covariance operator

$$\Sigma = \begin{pmatrix} \Sigma_u + Q_v & Q_v \\ Q_v & Q_v \end{pmatrix}, \tag{32}$$

where,

$$Q_v := A^\dagger \Sigma_v (A^\dagger)^*. \tag{33}$$

Lemma

The linear operator Q_v is trace class.

Moreover, the linear operator

$$T := Q_v \left[\Sigma_u + Q_v\right]^{-1/2}$$

is Hilbert-Schmidt.

Theorem

Under Assumptions (7), (8) and (9), the smoothing operator

$$\hat{B} := (A^{\dagger})^{*}\Sigma_u A^{*}\Sigma_v^{-1} \tag{34}$$

is the unique operator which satisfies

$$\hat{B} = \arg\min_{B} \|E[y|x] - y(B,x)\|_{H_1} ,$$

where the minimum is taken with respect to all linear closed and densely defined operators which satisfy the positivity condition (30).

Assuming that $u \sim N(0, \Sigma_u)$ and $v \sim N(0, \Sigma_v)$ where Σ_u and Σ_v are self-adjoint positive-definite bounded but not trace class operators on H_1 and H_2, respectively. In view of Assumption (7), the operator $A^\dagger : H_2 \to H_1$ is linear and

bounded operator. Put $H_3 := \operatorname{Ran} A$, H_3 is a Hilbert space, since it is a closed subspace of Hilbert space H_2. Let $\bar{A}^\dagger$ be the restriction of $A^\dagger$ on H_3 i.e. $\bar{A}^\dagger : H_3 \to H_1$. Hence $\bar{A}^\dagger$ is injective bounded linear operator.

Remark

In view of Hodrick-Prescott Filter (6), $v \in \operatorname{Ran}(A) = H_3$ i.e. it can be seen as H_3-random variable with covariance operator $\Sigma_v : H_3 \to H_3$.

Set

$$K_1 := (\bar{A}^\dagger(\bar{A}^\dagger)^*)^{-1} : H_1 \to H_1,$$

and

$$K_2 := ((\bar{A}^\dagger)^*\bar{A}^\dagger)^{-1} : H_3 \to H_3.$$

Assumption (10): There is $n_0 > 0$ such that the covariance operators $\tilde{\Sigma}_u, \tilde{\Sigma}$ and $\tilde{\Sigma}_v$ are trace class on the Hilbert spaces H_1^{-n} and H_3^{-n}, respectively, where

$$\tilde{\Sigma}_u = (\bar{A}^\dagger(\bar{A}^\dagger)^*)^n \Sigma_u (\bar{A}^\dagger(\bar{A}^\dagger)^*)^n, \quad \tilde{\Sigma}_v = ((\bar{A}^\dagger)^*\bar{A}^\dagger)^n \Sigma_v ((\bar{A}^\dagger)^*\bar{A}^\dagger)^n \tag{35}$$

and

$$\tilde{\Sigma} = \begin{pmatrix} \widetilde{\Sigma}_u + \widetilde{Q}_v & \widetilde{Q}_v \\ \widetilde{Q}_v & \widetilde{Q}_v \end{pmatrix}, \tag{36}$$

where,

$$\widetilde{Q}_v := \bar{A}^\dagger \widetilde{\Sigma}_v (\bar{A}^\dagger)^*. \tag{37}$$

Theorem

Let assumption 5 hold. Then, the unique optimal smoothing operator associated with the HP filter associated with H_1^{-n}-valued data x is given by:

$$\hat{B}h := (\bar{A}^\dagger)^* \tilde{\Sigma}_u A^* \tilde{\Sigma}_v^{-1} h, \quad h \in H_3^{-n}. \tag{38}$$

Assuming u and v independent and Gaussian random variables with zero means and covariance operators $\Sigma_u = \sigma_u I_{H_1}$ and $\Sigma_v = \sigma_v I_{H_3}$, where I_{H_1} and I_{H_3} denote the H_1 and H_3 identity operators, respectively and σ_u and σ_v are constant scalars. Assumption 5 reduces to

Assumption 11. There is an $n_0 > 0$ such that $\left(\bar{A}^{\dagger}(\bar{A}^{\dagger})^{*}\right)^{2n}$ and $\left((\bar{A}^{\dagger})^{*}\bar{A}^{\dagger}\right)^{2n}$ are trace class for all $n \geq n_0$.

Under this assumption, the associated covariance operators

$$\tilde{\Sigma}_u = (\bar{A}^\dagger(\bar{A}^\dagger)^*)^n \Sigma_u (\bar{A}^\dagger(\bar{A}^\dagger)^*)^n = \sigma_u \left(\bar{A}^\dagger(\bar{A}^\dagger)^*\right)^{2n},$$
$$\tilde{\Sigma}_v = ((\bar{A}^\dagger)^*\bar{A}^\dagger)^n \Sigma_v ((\bar{A}^\dagger)^*\bar{A}^\dagger)^n = \sigma_v \left((\bar{A}^\dagger)^*\bar{A}^\dagger\right)^{2n}$$

and

$$\widetilde{Q}_v = \sigma_v A^\dagger \left((A^\dagger)^* A^\dagger\right)^{2n} (A^\dagger)^* = \sigma_v \left(A^\dagger(A^\dagger)^*\right)^{2n+1}$$

are trace class, the expression (38) giving the optimal smoothing operator $\hat{B}$ reduces to

$$\hat{B} = (\bar{A}^\dagger)^* \tilde{\Sigma}_u A^* \tilde{\Sigma}_v^{-1} h = \frac{\sigma_u}{\sigma_v} I_{H_3^{-n}}, \tag{39}$$

i.e. $\hat{B}$ is the noise-to-signal ratio which is in the same pattern as in the classical HP filter.

- Agapiou, S., Larsson, S., Stuart A. M. (2012): Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems. *arXiv:1203.5753v2 [math.ST].*
- Bosq, D. (2000): Linear Processes in Function Spaces. *Lecture Notes in Statistics, Vol. 149, Springer-Verlag, New York.*
- Dermoune, A., Djehiche, B. and Rahmania, N. (2008): Consistent Estimator of the Smoothing Parameter in the Hodrick-Prescott Filter. *J. JapanStatist. Soc., Vol. 38 (No. 2), pp.225-241.*
- Dermoune, A., Djehiche, B. and Rahmania, N. (2009): Multivariate Extension of the Hodrick-Prescott Filter-Optimality and Characterization. *Studies in Nonlinear Dynamics & Econometrics, 13 , pp. 1-33.*
- Engl, H. W., Hanke, M., Neubauer, A. (1996): Regularization of inverse problems. Mathematics and its Applications, Vol. 375. Kluwer Academic Publishers, Dordrecht.
- Ferraty, F. and Vieu, P. (2006): Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations, Springer-Verlag, London, 2006.

- Frank, I. and Friedman, J. (1993): A statistical view of some chemometrics regression tools, *Technometrics 35 pp. 109-148.*
- Hastie, T., Buja, A., Tibshirani, R. (1995): Penalized discriminant analysis. *Ann. Statist. (23) pp. 73-102.*
- Hida T., (1980): Brownian Motion. *Springer-Verlag, New York-Berlin.*
- 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.
- Kaipio, J. and Somersalo, E. (2004): Statistical and Computational Inverse Problems. *Applied Mathematical Series, Vol. 160 Springer, Berlin.*
- Kuo, H.-H. (1975): Gaussian Measures in Banach Spaces. *Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin.*
- Lehtinen, M.S., Päivärinta, L. and Somersalo, E. (1989): Linear inverse problems for generalized random variables. Inverse Problems, 5:599-612.
- Mandelbaum, A. (1984): Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrscheinlichkeitstheorie Verw.*

- Müller, H.-G. and Stadtmüller, U. (2005): Generalized functional linear models, *The Annals of Statistics (33) pp. 774-805.*
- Preda, C. and Saporta, G. (2005): PLS regression on a stochastic process, *Computational Statistics & Data Analysis (48) pp. 149-158.*
- Ramsay, J. O. and Silverman, B. W. (1997): *Functional Data Analysis*, Springer-Verlag, New York.
- Rozanov, Ju. A. (1968): *Infinite-dimensional Gaussian distribution*. Proc. Steklov Inst. Math. (108) (Engl. transl. 1971 (Providence RI: AMS)).
- Schlicht, E. (2005): Estimating the Smoothing Parameter in the So-Called Hodrick-Prescott Filter. *J. Japan Statist. Soc., Vol. 35 No. 1, 99-119.*
- Skorohod A. V. (1974): *Integration in Hilbert Spaces*, Springer-Verlag, Berlin.
- Stuart, A.M. (2010): Inverse problems: a Bayesian perspective, *Acta Numerica 19 (2010), 451-559.*

Thanks for your attention!