Hiba Nassar

Functional Hodrick-Prescott Filter

Licentiate Thesis

Mathematics

2013
A thesis for the Degree of Licentiate of Philosophy in Mathematics.

Functional Hodrick-Prescott Filter
Hiba Nassar

Linnaeus University
School of Computer Science, Physics and Mathematics
SE-351 95 Växjö, Sweden
http://www.lnu.se
To Rani, and my parents
Abstract

The study of functional data analysis is motivated by their applications in various fields of statistical estimation and statistical inverse problems. In this thesis we propose a functional Hodrick-Prescott filter. This filter is applied to functional data which take values in an infinite dimensional separable Hilbert space. The filter depends on a smoothing parameter. In this study we characterize the associated optimal smoothing parameter when the underlying distribution of the data is Gaussian. Furthermore we extend this characterization to the case when the underlying distribution of the data is white noise.

Keywords: Inverse problems, adaptive estimation, Hodrick-Prescott filter, smoothing, trend extraction, Gaussian measures on a Hilbert space.
Acknowledgments

I would like to express my gratitude to all those who gave me the possibility to complete this licentiate. I want to thank the Department of Mathematics at Linnaeus University who gave me the opportunity to continue my study. Furthermore I am grateful to my supervisor, Docent Astrid Hilbert, whose generous guidance and support made it much interesting to work. I would like also to thank my second supervisor, Professor Boualem Djehiche, for giving his time and kind advice. He made it possible for me to work on a topic that is of great interest to me.

Many friends have helped me stay sane through these two and a half years. Their support and care helped me overcome setbacks and stay focused on my study. I would like to express my special thanks to Haidar Al-Talibi.

I am also grateful to my family in Syria whose love, support and warmth still reach me even though they are far away. My sincere thanks to my beloved husband, Rani Basna, provided the love, support, encouragement and help that kept me going.
# Contents

Abstract v

Acknowledgments vi

1 Theoretical background 1
   1.1 Singular value decomposition of a compact operator. . . . . . . 1
   1.2 The Moore-Penrose Pseudo Inverse. . . . . . . . . . . . . . 2
   1.3 Functional Data Analysis . . . . . . . . . . . . . . . . . . 3
   1.4 Hilbert Space-Valued Gaussian Random Variable. . . . . . . . 3
   1.5 Hilbert Scales. . . . . . . . . . . . . . . . . . . . . . . . . 4
   1.6 Generalized Gaussian Random Variables. . . . . . . . . . . . . 5
   1.7 Hodrick-Prescott Filter History. . . . . . . . . . . . . . . . . 6
   References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2 Results 11

3 Papers 14
   3.1 Functional Hodrick-Prescott Filter . . . . . . . . . . . . . . . 16
Chapter 1

Theoretical background

In this chapter we shall introduce the main mathematical tools that are used in the thesis.

1.1 Singular value decomposition of a compact operator.

Let \( K : \mathcal{X} \to \mathcal{Y} \) be a compact operator, where \( \mathcal{X}, \mathcal{Y} \) are separable Hilbert spaces (i.e., Hilbert spaces with countable bases). Let \( K^* \) denote the adjoint operator of \( K \) (i.e. \( (Kx, y) = (x, K^*y) \) for all \( x \in \mathcal{X} \) and \( y \in \mathcal{Y} \)). A system \( (\lambda_n, e_n, d_n) \) with \( n \in \mathbb{N}, \lambda_n > 0, e_n \in \mathcal{X}, d_n \in \mathcal{Y} \), is called singular value decomposition (SVD) of a compact operator \( K \) if the following conditions hold:

The sequence \( \{\lambda_n\} \) consists of non-negative numbers such that \( \{\lambda_n^2\} \) are the non-zero eigenvalues of the self adjoint operator \( K^*K \) arranged in non-increasing order and repeated according to their multiplicity, the \( \{e_n, e_n \in \mathcal{X}\} \) are the eigenvectors of the operator \( K^*K \) corresponding to \( \{\lambda_n^2\} \), and the sequence \( \{d_n\} \) is defined in terms of \( \{e_n\} \) as

\[
d_n = \frac{Ke_n}{\|Ke_n\|}.
\]

The sequence \( \{d_n\} \) is a complete orthonormal system of eigenvectors of the operator \( KK^* \). Moreover,

\[
K e_n = \lambda_n d_n, \quad K^* d_n = \lambda_n e_n, \quad n \in \mathbb{N},
\]

and the following decompositions hold:

\[
Kx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle d_n, \quad x \in \mathcal{X},
\]

and

\[
K^* y = \sum_{n=1}^{\infty} \lambda_n \langle y, d_n \rangle e_n, \quad y \in \mathcal{Y}.
\]
Chapter 1. Theoretical background

The infinite series in these decompositions converge in the norm of the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively.

1.2 The Moore-Penrose Pseudo Inverse.

The Moore-Penrose pseudo-inverse is a general way to find the solution to the following linear equation

$$Kx = y,$$

where $K : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator.

Definition 1.2.1. • An element $x \in \mathcal{X}$ that minimizes the functional

$$J(x) = \|Kx - y\|^2$$

is called a least-squares solution of $Kx = y$.

• $x \in \mathcal{X}$ is called best-approximate solution of $Kx = y$ if $x$ is a least-squares solution of $Kx = y$ and

$$\|x\| = \inf\{\|z\| \text{ where } z \text{ is least-squares solution of } Kx = y\}$$

holds.

To this end we introduce the notations $\mathcal{D}(A)$ and $\ker(A)$ for the domain of definition of the operator $A$ and the kernel of $A$, respectively. Now the definition of Moore-Penrose pseudo-inverse is provided, see [1].

Definition 1.2.2. The Moore-Penrose pseudo-inverse of $K$ is the operator $K^\dagger : \mathcal{D}(K^\dagger) \to \ker(K)^\perp$, which associates with each $y$ the unique minimum least square solution of $K$, and satisfies all of the following four criteria:

• $KK^\dagger K = K$,

• $K^\dagger KK^\dagger = K^\dagger$,

• $(KK^\dagger)^* = KK^\dagger$,

• $(K^\dagger K)^* = K^\dagger K$.

If $K$ is a compact linear operator with SVD $(\lambda_n, e_n, d_n)$ and $y \in \mathcal{D}(K^\dagger)$, then the vector

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle y, d_n \rangle e_n$$

holds.
is well defined and is a least squares solution. Therefore
\[ x = K^\dagger y = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle y, d_n \rangle e_n. \]

Special case: Let \( y \in D(K^\dagger) \) then multiply both sides of the equation \( Kx = y \) by \( K^* \), it gives
\[ K^* Kx = K^* y. \]
If \( K^* K \) is invertible (which is the case in this thesis) then the Moore-Penrose pseudo-inverse of \( K \) is \( K^\dagger = (K^* K)^{-1} K^* \).

1.3 Functional Data Analysis

Functional data analysis is a new direction of Statistics that started in the early 1990s, but its foundation go back to much earlier times where the theory of operators in Hilbert space and functional analysis had achieved some developments.

A functional data is not a set of single observation but rather a set of measurements depending on a parameter set in a continuous way that are to be regarded as a single curve or image. This is due to the ability of modern instruments to perform tightly spaced measurements so that these data can be seen as samples of curves. In the most simple setting, the sample consists of \( N \) curves \( X_1(t), X_2(t), \ldots, X_N(t), t \in \tau \). The set \( \tau \) is typically an interval of a line.

However, in many different applications \( \tau \) might be a subset of plane, or sphere. In those cases, the data are surfaces over a region, or more general functions over some domain, hence the term functional data.

1.4 Hilbert Space-Valued Gaussian Random Variable.

Definition 1.4.1. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \( H \) a Hilbert space with inner product \( \langle \cdot, \cdot \rangle \). A measurable function
\[ X : \Omega \to H \]
is called a Hilbert space-valued random variable.

A random variable \( X \) taking values in a separable Hilbert space \( H \) is said to be Gaussian if for every \( y \in H \) the law of \( \langle y, X \rangle \) is Gaussian.
Gaussian random variables are well defined by their mean \( m = E(X) \in H \) and their covariance operator \( \Sigma : H \to H \) defined by
\[
\langle y, \Sigma z \rangle = E[(y, X - m)(X - m, z)].
\]
The mean and the covariance operator are uniquely determined for that Hilbert space-valued Gaussian random variable is denoted by \( N(m, \Sigma) \).

The covariance operator of a Gaussian random variable has special properties which can be characterized by the following lemma.

**Lemma 1.4.1.** Let \( X \) be a Gaussian random variable on a separable Hilbert space. Then the covariance operator \( \Sigma \) of \( X \) is self-adjoint, positive and trace-class.

For more details see e.g. [19], [11].

### 1.5 Hilbert Scales.

Suppose \( H \) is a Hilbert space with inner product \( \langle \cdot, \cdot \rangle_H \). Let \( L : H \to H \) be a densely defined, unbounded, self-adjoint, positive operator, so that \( L \) is closed and satisfies \( \langle Lx, y \rangle = \langle x, Ly \rangle \) for every \( x, y \in D(L) \) and there exits a positive constant \( \gamma \) such that \( \langle Lx, x \rangle \geq \gamma \|x\|^2 \) for all \( x \in D(L) \). Let
\[
\mathcal{M} = \bigcap_{k=0}^{\infty} L^k H = \bigcap_{k=0}^{\infty} D(L^k)
\]
be the set of all elements \( x \) for which all the powers of \( L \) are defined. It is proved that this set is a dense set in \( H \), see e.g. [5]. For all \( s \in \mathbb{R} \) and all \( u, v \in \mathcal{M} \), define the inner products
\[
\langle x, y \rangle_s := \langle L^s x, L^s y \rangle_H,
\]
and the corresponding norms
\[
\|x\|_s := \|L^s x\|_H.
\]
The Hilbert space \( H_s \) denotes the completion of \( H \) with respect to the norm \( \|\cdot\|_s \). The family \( (H_s)_{s \in \mathbb{R}} \) is called the Hilbert scale induced by \( L \).

A Hilbert scale \( (H_s)_{s \in \mathbb{R}} \) possesses the following properties (see [2], [5] and [18]) :

- \( H_s \) is compactly embedded in \( H_r \) if \( s \geq r \) and is dense there.

- If \( s \geq 0 \) then \( H_{-s} \) is the dual space of \( H_s \). \( H_{-s} \) is equipped with the following inner product:
\[
\langle x, y \rangle_{-s} := \langle L^{-s} x, L^{-s} y \rangle_H.
\]
Let \(-\infty < q < r < s < \infty\) and \(x \in H_s\). Then the interpolation inequality
\[
\|x\|_r \leq \|x\|_s^{\delta-r} \|x\|_q^{r-q}
\]
holds.

### 1.6 Generalized Gaussian Random Variables.

Let \(N\) be the intersection of separable Hilbert spaces \((H_s, \|\cdot\|_s)\) i.e.
\[
N = \bigcap_{k=0}^{\infty} H_k
\]
where \(H_{s+1} \subset H_s\) and \(N\) is dense in each \(H_s\). The topology of \(N\) is the projective limit topology. Let \(N^*\) be the dual of \(N\) equipped with the weak topology. The space \(N^*\) can be viewed as a union of the duals of \(H_n\) i.e.
\[
N^* = \bigcup_{k=0}^{\infty} H_{-k}.
\]

Note: The family \((H_s)_{s \in \mathbb{R}}\) can be chosen as a Hilbert scale.

**Definition 1.6.1.** Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. A continuous mapping
\[
\eta : N \to L^2(\Omega, \mathcal{F}, \mathbb{P})
\]
is called a generalized Gaussian random variable if the random variables \(u \mapsto \eta(u)\) are Gaussian for all \(u \in N\).

The mean value and the covariance operator admit the following representations, respectively:
\[
E\eta(u) = m(u) = \langle u, m \rangle, \quad m \in N^*,
\]
\[
E((\eta(u) - m(u))(\eta(v) - m(v))) = \langle u, \Sigma v \rangle, \quad \Sigma : N \to N^*.
\]

**Proposition 1.6.1.** Let \(\eta\) be a generalized Gaussian random variable. Then there exists \(n_0 \in \mathbb{N}\) such that \(\eta\) belongs to \(H_{-n} \subset N^*\) with probability one for all \(n \geq n_0\). The covariance operator \(\Sigma\) extends continuously to a nuclear mapping
\[
\Sigma : H_n \to H_{-n}, \quad n \geq n_0.
\]

For more details see [15] and [19].
1.7 Hodrick-Prescott Filter History.

In this chapter we give a brief history of the Hodrick-Prescott filter. The procedure for filtering a trend as a smooth curve from claims data has a long history in actuarial science. The real starting point of this filter was proposed by Leser (1961), building on the graduation method developed by Whittaker (1923) and Henderson (1924), later on this filter is introduced into economics by Hodrick and Prescott in 1980 and 1997. Since that time the Hodrick-Prescott filter is widely used in economics and finance to estimate and predict business cycles and trends in macroeconomic time series. For a comparison of different filters as regards their applications in data analysis see [16] and references therein.

Hodrick and Prescott define a trend \( y = (y_1, y_2, ..., y_T) \) of a time series \( x = (x_1, x_2, ..., x_T) \) as the minimizer of

\[
\sum_{t=1}^{T} (x_t - y_t)^2 + \alpha \sum_{t=3}^{T} (y_t - 2y_{t-1} + y_{t-2})^2, \tag{1.1}
\]

with respect to \( y \), for an appropriately chosen positive parameter \( \alpha \), called the smoothing parameter. The first term in (1.1) measures a goodness-of-fit and the second term is a measure of the degree-of-smoothness which penalizes decelerations in growth rate of the trend component.

The program eViews recommends \( \alpha = 1600 \) when the time series \( x \) represents the log of U.S. real GNP. This value is also widely popular in the analysis of quarterly macroeconomic data.

Assuming the components of the residual (noise) \( u = x - y \) and the components of the signal \( v \) defined by

\[ v_t = y_t - 2y_{t-1} + y_{t-2} := D^2 y_t, \quad t = 3, 4, \ldots T, \]

to be independent and Gaussian, where \( D^2 \) stands for the second order backwards shift operator. The operator \( D^2 \) can be written in vector form

\[ Py(t) = D^2 y_t, \quad t = 3, 4, \ldots, T, \]

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

\[
\begin{pmatrix}
1 & -2 & 1 & 0 & \ldots & \ldots & 0 \\
0 & 1 & -2 & 1 & \ldots & \ldots & 0 \\
0 & 0 & 1 & -2 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 1 & -2 & 1
\end{pmatrix}.
\]

Hence the time series of observations \( x \) satisfies the following linear mixed model:
1.7. Hodrick-Prescott Filter History.

\[ \begin{cases} x = y + u, \\ Py = v. \end{cases} \]  \hspace{1cm} (1.2)

where \( u \sim N(0, \sigma_u^2 I_T) \) and \( v \sim N(0, \sigma_v^2 I_{T-2}) \).

To find an explicit expression of the trend \( y(\alpha, x) \) which minimizes (1.1), we introduce the following well known result (see e.g. [4]):

**Proposition 1.7.1.** Let \( M_1, M_2 \) be two non-negative matrices. If \( M_1 + M_2 \) is invertible. Then, for each fixed \( x \), the minimizer

\[ \arg \min_y \left[ (x - y)' M_1 (x - y) + y' M_2 y \right] \]

is given by \( y(M_1, M_2, x) = (M_1 + M_2)^{-1} M_1 x. \)

Since \( P \) is of rank \( T - 2 \), the matrix \( PP' \) is invertible. Applying Proposition (1.7.1) we obtain the unique positive definite solution

\[ y(\alpha, x) = (I_T + \alpha P' P)^{-1} x. \]  \hspace{1cm} (1.3)

Hodrick and Prescott (1980) and later Schlicht (2005) introduced a way to estimate the smoothing parameter \( \alpha \). The optimal \( \alpha \) is the one which let the optimal solution \( y(\alpha, x) \) in (1.3) be the best predictor of \( y \) given the time series \( x \) i.e.

\[ y(\alpha, x) = E[y|x]. \]  \hspace{1cm} (1.4)

In order to estimate the smoothing parameter, obviously the joint distribution of \( (x, y) \) is needed. That explains why Hodrick and Prescott assume the Gaussian distributions for both the noise and the signal. Moreover, in this case the conditional expectation can be computed explicitly which is not always the case.

Schlicht (2005) gave a characterization of the optimal smoothing parameter \( \alpha \) as so-called noise-to-signal ratio i.e. the ratio of the variance of the noise to the variance of the signal.

Depending on the criterion (1.4), Dermoune et al. (2007) derived a consistent estimator of the noise-to-signal ratio \( \alpha \), and non-asymptotic confidence intervals for \( \alpha \) given a confidence level.

Furthermore Dermoune et al. (2009) suggest an optimality criterion for choosing the smoothing parameter for the HP-filter. The smoothing parameter \( \alpha \) is chosen to be the one which minimizes the difference between the optimal solution \( y(\alpha, x) \) and the best predictor of trend \( y \) given the time series \( x \) i.e.

\[ \alpha^* = \arg \min_{\alpha > 0} \| E[y|x] - y(\alpha, x) \|^2. \]  \hspace{1cm} (1.5)
Moreover, the optimality of this smoothing parameter has been studied two different cases:

- under singular-value parametrization of the trend.
- under the initial-value parametrization of the trend.

In addition, Dermoune et al. suggested an extension to the HP filter to the multivariate case, where they show that the possible optimal smoothing parameters that satisfy (1.5) constitute a whole set of pairs of positive definite matrices by using singular-value parametrization of the trend.
Bibliography


Bibliography


Chapter 2

Results

In this chapter we present a summary of the results of this thesis. In this thesis an extension of Hodrick-Prescott filter is proposed in the case where we have functional data. The functional Hodrick-Prescott filter is defined as a mixed model of the same form as (1.2) where the operator $P$ is replaced by a compact operator $A$ between two Hilbert spaces:

\[
\begin{aligned}
  x &= y + u, \\
  Ay &= v,
\end{aligned}
\]  \hspace{1cm} (2.1)

where $x$ is a functional time series of observables, and the noise, $u$, as well as the signal, $v$, are Gaussian with trace class covariance operators, i.e. $u \sim N(0, \Sigma_u)$ and $v \sim N(0, \Sigma_v)$. The optimal smooth trend associated with $x$ satisfies

\[\|x - y\|_{H_1}^2 + \langle Ay, BAy \rangle_{H_2}\]  \hspace{1cm} (2.2)

with respect to $y$, where $B$ is a regularization smoothing operator instead of the smoothing parameter $\alpha$ in (1.1).

The main result of this study is the characterization of the optimal smoothing operator $\hat{B}$ which can be formalized in the following theorem, which is the infinite dimensional extension of what Dermoune et al. did in (2009).

**Theorem 2.0.2.** Under some necessary assumptions, for all $x \in H_1$, the smoothing operator (which is linear, compact and injective)

\[\hat{B}h := (AA^*)^{-1}A\Sigma_u A^*\Sigma_v^{-1}h,\]  \hspace{1cm} (2.3)

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 a positivity condition.

Furthermore, we have

\[y(\hat{B}, x) - E[y|x] = (I_{H_1} - \Pi)(x - E[x]),\]  \hspace{1cm} (2.4)
Chapter 2. Results

and its covariance operator is

\[
\text{cov}(y(\hat{B},x) - E[y|x]) = (I_{H_1} - \Pi)\Sigma_u. \tag{2.5}
\]

In particular,

\[
E\left(\|y(\hat{B},x) - E[y|x]\|_{H_1}^2\right) = \text{trace}\left((I_{H_1} - \Pi)\Sigma_u\right), \tag{2.6}
\]

where \(\Pi := A^* (AA^*)^{-1} A\).

The second main result is to characterize the optimal smoothing operator \(\hat{B}\) when the covariance operators of the signal and the noise are non-trace class, this includes the important case where the signal and the noise are white noise. Looking at these Gaussian variables as generalized random variables and under some assumptions to find an appropriate Hilbert scale where the covariance operators can be maximally extended to trace class operators on an appropriate domain. An explicit expression of the smoothing operator \(\hat{B}\) is given in a form similar to (2.3) but in the appropriate domain.

The third result in this thesis is to show that the optimal smoothing operator \(\hat{B}\) reduces to the noise-to-signal ratio where the noise and the signal are white noise which coincide with the previous studies.
Chapter 3

Papers
3.1 Functional Hodrick-Prescott Filter

Boualem Djehiche and Hiba Nassar