



Bloechl, Andreas:

Penalized Splines, Mixed Models and the Wiener-Kolmogorov Filter

Munich Discussion Paper No. 2014-44

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at <https://doi.org/10.5282/ubm/epub.21406>

Penalized Splines, Mixed Models and the Wiener-Kolmogorov Filter

Discussion Paper

Abstract

Penalized splines are widespread tools for the estimation of trend and cycle, since they allow a data driven estimation of the penalization parameter by the incorporation into a linear mixed model. Based on the equivalence of penalized splines and the Hodrick-Prescott filter, this paper connects the mixed model framework of penalized splines to the Wiener-Kolmogorov filter. In the case that trend and cycle are described by ARIMA-processes, this filter yields the mean squared error minimizing estimations of both components. It is shown that for certain settings of the parameters, a penalized spline within the mixed model framework is equal to the Wiener-Kolmogorov filter for a second fold integrated random walk as the trend and a stationary ARMA-process as the cyclical component.

JEL-Code: C220, C520

Keywords: Hodrick-Prescott filter, mixed models, penalized splines, trend estimation, Wiener-Kolmogorov filter

Andreas Blöchl
University of Munich
Schackstraße 4
80539 Munich
andreas.bloechl@lrz.uni-muenchen.de

1 Introduction

There is a rising popularity of penalized splines as instruments for trend estimation in economics. They are closely related to the predominantly used Hodrick-Prescott filter (Hodrick/Prescott 1997), which traces back to the ideas of Whittaker (1923), Henderson (1924) and Leser (1961). The output of both methods depends on the choice of a penalization parameter λ that controls the smoothness of the estimated trend. For the Hodrick-Prescott filter the most commonly used value for λ traces back to Hodrick/Prescott (1997), who propose to use $\lambda = 1600$ for quarterly data. This is based on the equivalence of the Hodrick-Prescott filter to an optimal Wiener-Kolmogorov filter (Whittle 1983, Bell 1984) for a second fold integrated random walk as the trend and a white noise process as the cyclical component, as well as subjective assumptions about the variance of the growth rates of trend and cycle. As these assumptions are rather restrictive, the suggestion of Hodrick/Prescott is criticized as dubious (Danthine/Girardin 1989). Furthermore, it is criticized as not data driven (e.g. Schlicht 2005, Kauermann et al. 2011), especially since the choice of Hodrick/Prescott refers to U.S. time series only.

To this point Schlicht (2005) shows how to write the Hodrick-Prescott filter as a linear mixed model in order to estimate the penalization parameter. However, this approach is limited to a white noise residual structure and thus a white noise business cycle, what is seldom true for economic time series. Nevertheless, this shortcoming can be corrected, as the Hodrick-Prescott filter belongs to the class of penalized splines (Paige 2010). Penalized splines offer the advantage to estimate trend and cycle data driven, where the most common methods are generalized cross validation (Hastie/Tibshirani 1990, also Ruppert 2002), or the incorporation into a mixed model framework (e.g. Brumback et al. 1999). Both methods allow for an autocorrelated residual structure (see Kohn et al. (1992) or Wang (1998) for the generalized cross validation with correlated errors). But while generalized cross validation is very sensitive about the specification of the autocorrelation structure (Opsomer et al. 2001, Proietti 2005), the mixed model approach appears to be relatively robust, as long as the deviation between assumed and true (but unknown) autocorrelation structure is not too large (Krivobokova/Kauermann 2007). It follows immediately from the equivalence of the Hodrick-Prescott filter and penalized splines that the Hodrick-Prescott filter can also be written as a linear mixed model with an autocorrelated error term. As a novel contribution this paper shows that in this case the optimal Wiener-Kolmogorov filter for a second fold integrated random walk as the trend and an autocorrelated cycle arises.

The first sections briefly summarize the Hodrick-Prescott filter and penalized splines with a truncated polynomial basis (Brumback et al. 1999) and describe the link between both methods. Moreover, it is explained how penalized splines can be interpreted as linear mixed models. Afterwards it is shown how these methods fit into the framework of the Wiener-Kolmogorov filter. Finally, the mixed model framework of the Hodrick-Prescott filter is used to estimate the trend component of the German industrial production index.

2 Model framework

2.1 The Hodrick-Prescott filter

The Hodrick-Prescott filter (henceforth denoted as HP-filter) decomposes a time series $\{y_t\}_{t=1}^T$ into two components, i.e. $y_t = \mu_t + c_t$. μ_t is regarded as the trend while c_t represents the rest, usually the sum of cycle and irregular effects. The trend μ_t is estimated by solving the following minimization problem:

$$\min_{\mu_t} \sum_{t=1}^T (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2. \quad (1)$$

The first part of the minimization problem causes a close fit of the estimation to the original data, while the second part penalizes the volatility of the trend. The parameter λ controls the smoothness of the trend function. Increasing λ makes the trend component become less flexible. For $\lambda = 0$ the trend is equal to the original series, whereas for $\lambda \rightarrow \infty$ it is a straight line. The solution to the minimization in (1) can be expressed in matrix notation (McElroy 2008, Faig 2012 p.16), which allows a fast and easy calculation:

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} + \lambda \boldsymbol{\Delta}' \boldsymbol{\Delta})^{-1} \mathbf{y}, \quad (2)$$

where $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_T)'$, $\mathbf{y} = (y_1, \dots, y_T)'$ and $\boldsymbol{\Delta} \in \mathbb{R}^{(T-2) \times T}$ is a differencing matrix, such that the product of $\boldsymbol{\Delta}$ and \mathbf{y} yields the second differences of \mathbf{y} . $\boldsymbol{\Delta}$ is described by formula (22) in section 3.

2.2 Penalized tp-splines

This section describes briefly penalized splines (O'Sullivan 1986, Eilers/Marx 1996). For a detailed discussion see Ruppert et al. (2003). Although there are many different types of splines, it is focused on truncated polynomial splines (Brumback et al. 1999, henceforth denoted as tp-splines). Even if tp-splines tend to be numerical instable, they are advantageous due to their relative easy implementation as well as their straight link to linear mixed models and the HP-filter. Estimating the trend component with a tp-spline means the trend function is modelled in dependence of time t . After dividing the variable t , $t = 1, \dots, T$, into $m - 1$ intervals by setting m knots $1 = \kappa_1 < \kappa_2 < \dots < \kappa_m = T$ a tp-spline of degree l for a time series $\{y_t\}_{t=1}^T$ can be written to:

$$y_t = f(t) + \varepsilon_t = \delta_1 + \delta_2 t + \dots + \delta_{l+1} t^l + \delta_{l+2} (t - \kappa_2)_+^l + \dots + \delta_d (t - \kappa_{m-1})_+^l + \varepsilon_t, \quad (3)$$

$$\text{with } (t - \kappa_j)_+^l = \begin{cases} (t - \kappa_j)^l & , t \geq \kappa_j \\ 0 & , \text{else} \end{cases},$$

where ε_t denotes the error term that represents the business cycle and $d = m + l - 1$. In matrix notation the model is defined to:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (4)$$

$$\text{with } \mathbf{Z} = \begin{pmatrix} 1 & 1 & \dots & 1^l & (1 - \kappa_2)_+^l & \dots & (1 - \kappa_{m-1})_+^l \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & T & \dots & T^l & (T - \kappa_2)_+^l & \dots & (T - \kappa_{m-1})_+^l \end{pmatrix}.$$

Here, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_T)'$ and $\mathbf{y} = (y_1, \dots, y_T)'$. tp-splines can be interpreted as a continuous function of piecewise defined polynomials of degree l . The coefficient of the highest polynomial changes at every knot because of the truncated polynomials, which allows a high flexibility of $f(t)$. Consequently, the smoothness of the trend can be determined by controlling the parameters of the truncated polynomials $\delta_{l+2}, \dots, \delta_d$, since they regulate the change of the coefficient of the highest polynomial (e.g. Fahrmeir et al. 2009). This can be done by estimating the vector of coefficients by minimizing the penalized least squares criterion.

$$PLS(\lambda) = \sum_{t=1}^T [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^d \delta_j^2. \quad (5)$$

The first part of $PLS(\lambda)$ aims at a close fit of the trend to the observed series, while the second part penalizes a too high volatility. This tradeoff is solved by the parameter λ that puts weight on the second part. Increasing the value for λ reduces the volatility of the trend. The solution of (5) is given to (e.g. Fahrmeir et al. 2009 p.313)

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K})^{-1}\mathbf{Z}'\mathbf{y}, \quad \text{where } \mathbf{K} = \text{diag}(\underbrace{0, \dots, 0}_{l+1}, \underbrace{1, \dots, 1}_{m-2}). \quad (6)$$

$$\text{so that } \hat{\mathbf{y}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K})^{-1}\mathbf{Z}'\mathbf{y}. \quad (7)$$

Given formula (7) it can be shown that for a certain selection for the parameters m and l the penalized tp-spline is equal to the HP-filter (Paige 2010). Setting $l = 1$ and knots at every point in time t , $t = 1, 2, \dots, T$, implies that the design matrix \mathbf{Z} is quadratic and invertible (c.f. Paige 2010 p.870). In this case it is generally defined as

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 4 & 2 & 1 & 0 & \dots & 0 & 0 \\ 1 & 5 & 3 & 2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & T-1 & T-3 & T-4 & T-5 & \dots & 1 & 0 \\ 1 & T & T-2 & T-3 & T-4 & \dots & 2 & 1 \end{pmatrix} \in \mathbb{R}^{T \times T}.$$

Given this special structure of \mathbf{Z} , formula (7) can also be expressed as (a derivation is provided in appendix A):

$$\hat{\mathbf{y}} = (\mathbf{I} + \lambda \mathbf{Z}'^{-1} \mathbf{K} \mathbf{Z}^{-1})^{-1} \mathbf{y}. \quad (8)$$

Taking into account that $\hat{\mathbf{y}} = \hat{\boldsymbol{\mu}}$ and that $\mathbf{Z}'^{-1} \mathbf{K} \mathbf{Z}^{-1} = \boldsymbol{\Delta}' \boldsymbol{\Delta}$ (a proof is given in appendix B), it becomes obvious that the HP-filter is identical to a penalized tp-spline of order one and with knots at every observed point in time $t = 1, \dots, T$. Note that this formulation is also equivalent to the so called local linear model (Proietti 2007). The equivalence between tp-splines and the HP-filter allows a useful interpretation of the HP-filter. The trend generated by the HP-filter is a continuous connection of lines that can change their slope at the points in time $t = 2, 3, \dots, T - 1$. The penalization parameter λ controls to what extent the slope of the lines can change at these points in time. Choosing high values for λ means the slope can change only slightly, which results in a smooth estimated trend. Furthermore, the equivalence of the HP-filter and a penalized spline implies that the mixed model framework with an autocorrelated residual structure can also be applied to the HP-filter to derive a data driven estimation for its penalization parameter.

2.3 Splines within a mixed model framework

Penalized tp-splines can be interpreted as a linear mixed model in order to derive a data driven estimation of λ (for a detailed discussion of mixed models see Searle et al. 1992, Vonesh/Chinchilli 1997, Pinheiro/Bates 2000 or McCulloch/Searle 2001). A tp-spline within a mixed model framework is defined to (e.g. Ruppert et al. 2003 p.108)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (9)$$

$$\text{where } \mathbf{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2^l \\ \vdots & \vdots & \ddots & \vdots \\ 1 & T & \dots & T^l \end{pmatrix} \text{ and } \mathbf{U} = \begin{pmatrix} (1 - \kappa_2)_+^l & \dots & (1 - \kappa_{m-1})_+^l \\ \vdots & \ddots & \vdots \\ (T - \kappa_2)_+^l & \dots & (T - \kappa_{m-1})_+^l \end{pmatrix}.$$

Consequently, $\mathbf{Z} = [\mathbf{X}, \mathbf{U}]$, $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \boldsymbol{\gamma}']$, $\boldsymbol{\beta} \in \mathbb{R}^{(l+1) \times 1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{(m-2) \times 1}$. Moreover, it is assumed that $\boldsymbol{\varepsilon} \sim \text{N}(0, \mathbf{R})$ and $\boldsymbol{\gamma} \sim \text{N}(0, \mathbf{G})$, so that in general an autocorrelated residual structure can be allowed. When the tp-spline is written as a mixed model, it can equivalently be interpreted as a hierarchical model (e.g. Fahrmeir et al. 2009):

$$\mathbf{y} | \boldsymbol{\gamma} \sim \text{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma}, \mathbf{R}) \text{ and the marginal distribution } \boldsymbol{\gamma} \sim \text{N}(0, \mathbf{G}).$$

Furthermore, the spline can be described by a marginal model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \text{ with } \boldsymbol{\varepsilon}^* = \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

The distribution of \mathbf{y} is then defined by $\mathbf{y} \sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, where $\mathbf{V} = \mathbf{R} + \mathbf{U}\mathbf{G}\mathbf{U}'$. Fur-

thermore, it is assumed that $\mathbf{R} = \sigma^2\mathbf{\Omega}$ and $\mathbf{G} = \text{diag}(\tau^2)$ (e.g. Kauermann et al. 2011), i.e. the error term is autocorrelated with a constant variance σ^2 and the autocorrelation matrix $\mathbf{\Omega}$ and the parameters $\boldsymbol{\gamma}$ are uncorrelated and have the constant variance τ^2 . Given the hierarchical model of \mathbf{y} the parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the resulting log-likelihood function with respect to $\boldsymbol{\theta}$ (e.g. Ruppert et al. 2003):

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\gamma}'\mathbf{G}^{-1}\boldsymbol{\gamma}. \quad (10)$$

This is equivalent to minimizing (e.g. Fahrmeir et al. 2009, Kauermann et al. 2011)

$$\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})'\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}) + \lambda\boldsymbol{\gamma}'\boldsymbol{\gamma}, \quad (11)$$

$$\text{where } \lambda = \frac{\sigma^2}{\tau^2}.$$

As $\mathbf{G} = \text{diag}(\tau^2)$, the solution of the maximization of formula (10) is (e.g. Robinson 1991, Hayes/Haslett 1999, also Ruppert et al. 2003)

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\gamma}} \end{bmatrix} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \frac{1}{\tau^2}\mathbf{K})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}, \quad (12)$$

where the estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ are the best linear unbiased predictors (BLUPs) of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

In most situations the covariance matrices \mathbf{R} and \mathbf{G} are unknown and have to be estimated. To this regard both matrices are written in dependence of a vector of parameters $\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}$ depends on the assumed autocorrelation structure of $\boldsymbol{\varepsilon}$. From this the log-likelihood is derived by the interpretation as a marginal model. It is given except of a constant term to (e.g. Ruppert et al. 2003):

$$l(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = -\frac{1}{2} [\log(|\mathbf{V}(\boldsymbol{\vartheta})|) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}(\boldsymbol{\vartheta})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]. \quad (13)$$

Differentiating and solving with respect to $\boldsymbol{\beta}$ yields

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = [\mathbf{X}'\mathbf{V}(\boldsymbol{\vartheta})^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}(\boldsymbol{\vartheta})^{-1}\mathbf{y}. \quad (14)$$

Reinserting into (13) yields the profile log-likelihood:

$$l_p(\boldsymbol{\vartheta}) = -\frac{1}{2} \left(\log |\mathbf{V}(\boldsymbol{\vartheta})| + [\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})]'\mathbf{V}(\boldsymbol{\vartheta})^{-1}[\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})] \right). \quad (15)$$

Instead of the profile log-likelihood in most applications the restricted log-likelihood (e.g. Searle et al. 1992) is maximized:

$$l_r(\boldsymbol{\vartheta}) = l_p(\boldsymbol{\vartheta}) - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}(\boldsymbol{\vartheta})^{-1}\mathbf{X}|. \quad (16)$$

The restricted log-likelihood is more accurate in small samples, because it takes into account

the degrees of freedom for of the fixed effects (Ruppert et al. 2003). The maximization of $l_r(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ yields $\hat{\boldsymbol{\vartheta}}$ and thus the estimators $\hat{\mathbf{R}}$ and $\hat{\mathbf{G}}$. Because σ^2 and τ^2 are contained in $\boldsymbol{\vartheta}$ also $\hat{\lambda} = \frac{\hat{\sigma}^2}{\hat{\tau}^2}$ can be received. The estimators for the autocovariance matrix $\hat{\mathbf{R}}$ and $\hat{\tau}^2$ can be inserted into (12) which yields the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ that are called the estimated BLUPs or EBLUPs of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

An advantage of the estimation of tp-splines within the mixed model framework is that the results hardly depend on the number of knots. In general the flexibility of the spline increases when a higher number of knots is selected. However, Kauermann/Opsomer (2011) demonstrate that the mixed model framework compensates the higher number of knots by an increase of the penalization. As a consequence, the number of knots has no significant effect on the results, as long as it is not too low. Furthermore, Krivobokova/Kauermann (2007) demonstrate that the results are robust with regard to a misspecification of the residual autocorrelation structure. However, the assumed autocorrelation structure may not be too different from the true (but unknown) one.

3 Penalized splines and the Wiener-Kolmogorov filter

It is known that the HP-filter can be integrated into the framework of the Wiener-Kolmogorov filter (Hodrick/Prescott 1997, for the Wiener-Kolmogorov filter see Whittle 1983, Bell 1984 also Harvey 1989 and Kaiser/Maravall 2001). As penalized splines are equal to the HP-filter under certain values for the parameters (Proietti/Luati 2007, Paige 2010) there is also a link between penalized splines and the Wiener-Kolmogorov filter. To understand the relationship between these filters, it is useful to derive their connection from a very general standpoint. Assume that a time series $\{y_t\}_{t=1}^T$ can be written as the sum of a trend μ_t and a cyclical component c_t

$$y_t = \mu_t + c_t. \quad (17)$$

This can equivalently be expressed in matrix notation:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{c}, \quad (18)$$

where $\mathbf{y} = (y_1, \dots, y_T)'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$ and $\mathbf{c} = (c_1, \dots, c_T)'$. If trend and cycle are defined as ARIMA(p, d, q) models and L denotes the backshift operator such that $L^k y_t = y_{t-k}$, then the components can be expressed as:

$$\Phi^\mu(L)\mu_t = \Theta^\mu(L)\varepsilon_t \quad \text{and} \quad \Phi^c(L)c_t = \Theta^c(L)\eta_t, \quad (19)$$

$$\text{where } \Phi^\mu(L) = 1 - \varphi_{\mu,1}L^1 - \dots - \varphi_{\mu,p}L^p,$$

$$\text{and } \Theta^\mu(L) = 1 + \theta_{\mu,1}L^1 + \dots + \theta_{\mu,q}L^q.$$

$\Phi^c(L)$ and $\Theta^c(L)$ are defined analogously. Both ε_t and η_t are independent white noise variables. If μ_t and c_t are integrated of (arbitrary) order n and r , $\Phi^\mu(L)$ and $\Phi^c(L)$ have char-

acteristic polynomials with n and r roots on the unit circle. Consequently, $(1-L)^n \mu_t = u_t$ and $(1-L)^r c_t = v_t$ are stationary and the vectors $\mathbf{u} = (u_{n+1}, \dots, u_T)'$ and $\mathbf{v} = (v_{r+1}, \dots, v_T)'$ have the autocovariance matrices $\Sigma_{\mathbf{u}}$ and $\Sigma_{\mathbf{v}}$. By introducing differencing matrices Δ_{μ} and Δ_c of dimension $(T-n) \times T$ and $(T-r) \times T$, where as an example the product of Δ_{μ} and μ yields the n^{th} first differences of μ , this can be expressed as:

$$\Delta_{\mu} \mu = \mathbf{u} \quad \text{and} \quad \Delta_c \mathbf{c} = \mathbf{v}.$$

On purpose to estimate the trend component μ McElroy (2008) shows that the minimum mean squared error linear estimate of μ is given by:

$$\hat{\mu} = (\Delta_c' \Sigma_{\mathbf{v}}^{-1} \Delta_c + \Delta_{\mu}' \Sigma_{\mathbf{u}}^{-1} \Delta_{\mu})^{-1} \Delta_c' \Sigma_{\mathbf{v}}^{-1} \Delta_c \mathbf{y}. \quad (20)$$

This is the matrix formulation of the Wiener-Kolmogorov filter. The autocovariance matrices of \mathbf{u} and \mathbf{v} can equivalently be written as $\Sigma_{\mathbf{u}} = \Gamma_{\mathbf{u}} \sigma_u^2$ and $\Sigma_{\mathbf{v}} = \Gamma_{\mathbf{v}} \sigma_v^2$, where $\Gamma_{\mathbf{u}}$ and $\Gamma_{\mathbf{v}}$ are the autocorrelation matrices of \mathbf{u} and \mathbf{v} , while σ_u^2 and σ_v^2 are the respective variances. Consequently, $\hat{\mu}$ can also be expressed by:

$$\hat{\mu} = (\Delta_c' \Gamma_{\mathbf{v}}^{-1} \Delta_c + \frac{\sigma_v^2}{\sigma_u^2} \Delta_{\mu}' \Gamma_{\mathbf{u}}^{-1} \Delta_{\mu})^{-1} \Delta_c' \Gamma_{\mathbf{v}}^{-1} \Delta_c \mathbf{y}. \quad (21)$$

Now, consider the special case, where the trend is a second order integrated random walk and the cyclical component is a white noise process. Then the model above simplifies to:

$$(1-L)^2 \mu_t = \varepsilon_t,$$

$$c_t = \eta_t.$$

This implies that $\Gamma_{\mathbf{u}} = \Gamma_{\mathbf{v}} = \mathbf{I}$ as well as $\Delta_c = \mathbf{I}$. The differencing matrix $\Delta_{\mu} \in \mathbb{R}^{(T-2) \times T}$ is defined as:

$$\Delta_{\mu} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}. \quad (22)$$

Thus, the minimum mean squared error estimation of the trend component reduces to (Flaig 2012 p.16):

$$\hat{\mu} = (\mathbf{I} + \frac{\sigma_v^2}{\sigma_u^2} \Delta_{\mu}' \Delta_{\mu})^{-1} \mathbf{y}, \quad (23)$$

what is the matrix formula for the Hodrick-Prescott filter, when λ is equal to the inverse signal to noise ratio $\frac{\sigma_v^2}{\sigma_u^2}$. This makes obvious that the HP-filter is minimizing the mean squared error, when the trend is a two-fold integrated random walk and the cycle is just white noise. From this Hodrick/Prescott (1997) suggest to set $\sigma_v = 5$ and $\sigma_u = 1/8$ for quarterly data, which results in a penalization of $\lambda = 1600$, that has meanwhile become an "industry standard" in economics (Flaig 2012 p.23).

When the HP-filter is written as a penalized spline and incorporated into a mixed model, then usually an autocorrelated residual structure \mathbf{R} is assumed. It can be shown that the formulation of the HP-filter as a mixed model according to (12) is a special case of the Wiener-Kolmogorov filter in (20) or (21). Assume that the trend component is just like before a second order integrated random walk, but the cycle is now an autocorrelated stationary process with the autocorrelation structure $\mathbf{\Gamma}_v$ and variance σ_v^2 . Then it follows that $\mathbf{\Gamma}_u = \mathbf{I}$ and $\mathbf{\Delta}_c = \mathbf{I}$, while $\mathbf{\Delta}_\mu$ is still defined like in (22). For these assumptions the Wiener-Kolmogorov filter is defined as

$$\hat{\boldsymbol{\mu}} = (\mathbf{\Gamma}_v^{-1} + \frac{\sigma_v^2}{\sigma_u^2} \mathbf{\Delta}_\mu' \mathbf{\Delta}_\mu)^{-1} \mathbf{\Gamma}_v^{-1} \mathbf{y}. \quad (24)$$

When the setting of the tp-spline is such that it is equal to the HP-filter, i.e. T equidistant knots are chosen and $l = 1$ is selected, then \mathbf{Z} is generally defined like in section 2.2. Using this, the mixed model representation of the HP-filter from (12) can be written as (the derivation is equivalent to appendix A).

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \frac{1}{\tau^2}\mathbf{K})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} = \mathbf{Z}(\mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{Z} + \frac{\sigma^2}{\tau^2}\mathbf{K})^{-1}\mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{y} = \\ &= (\mathbf{\Omega}^{-1} + \frac{\sigma^2}{\tau^2}\mathbf{Z}'^{-1}\mathbf{K}\mathbf{Z}^{-1})^{-1}\mathbf{\Omega}^{-1}\mathbf{y}. \end{aligned} \quad (25)$$

Noting that $\mathbf{Z}'^{-1}\mathbf{K}\mathbf{Z}^{-1} = \mathbf{\Delta}_\mu' \mathbf{\Delta}_\mu$ then it immediately follows that

$$\hat{\mathbf{y}} = (\mathbf{\Omega}^{-1} + \frac{\sigma^2}{\tau^2} \mathbf{\Delta}_\mu' \mathbf{\Delta}_\mu)^{-1} \mathbf{\Omega}^{-1} \mathbf{y}. \quad (26)$$

As $\mathbf{\Omega} = \mathbf{\Gamma}_v$ and $\lambda = \frac{\sigma^2}{\tau^2} = \frac{\sigma_v^2}{\sigma_u^2}$ it becomes obvious that the HP-filter within the mixed model framework is equal to the Wiener-Kolmogorov filter when the trend is a two-fold integrated random walk and the cycle is a stationary autocorrelated process. It follows for this setting that the Wiener-Kolmogorov filter can be estimated by maximum likelihood estimation. Moreover, in this setting a penalized tp-spline within a mixed model is the mean squared error minimizing filter for T equidistant knots and $l = 1$.

4 Empirical application

In this section the HP-filter is used to estimate the trend component of the German seasonally adjusted real industrial production index from January 1991 to October 2013.¹ The HP-filter is written as a penalized tp-spline within a mixed model framework in order to derive an estimator for the penalization parameter λ . Although this setting is special, the results can be seen as general for penalized splines and the mixed model interpretation. Kauermann/Opsomer (2011) show that the results are almost independent of the number

¹The data of the industrial production are from the German Federal Statistical Office.

of knots m , as long as m is not too low. Furthermore, Claeskens et al. (2009) show that the estimations are hardly influenced by the selected spline basis, which is also demonstrated by Ruppert (2002).

An important question concerns the assumed autocorrelation structure of the error term. According to Krivobokova/Kauermann (2007) the results of the maximum likelihood estimation are robust to a misspecification of the autocorrelation structure, as long as the assumed autocorrelation doesn't deviate too much from the true autocorrelation structure. However, to check for the robustness of the results with regard to the assumed residual autocorrelation structure, the estimation is done for AR(1) and AR(4) as well as ARMA(1,1) and ARMA(4,4) residual structures. The industrial production index contains short term fluctuations, i.e. a noise component (Flaig 2005 p.419). Thus, the assumption of an ARMA-process appears to be straightforward, since the sum of an AR(p)-process and a white noise variable yields an ARMA(p, p)-process (e.g. Schlittgen/Streitenberg 2001 p.133). The results of the estimation for the industrial production index are shown in Figure 1.

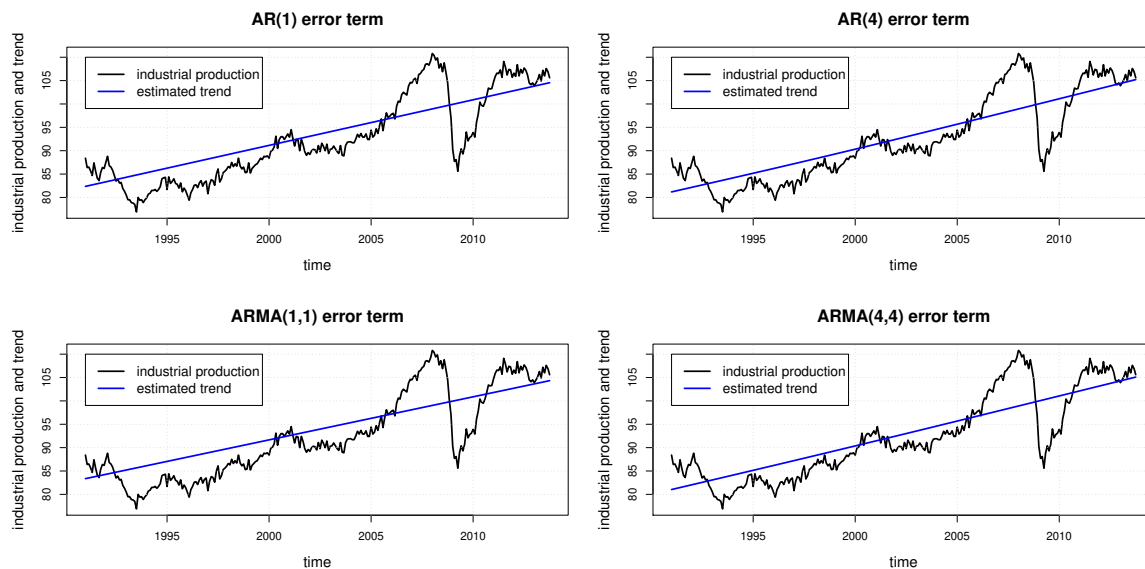


Fig. 1: Trend of the German industrial production index for different residual autocorrelation structures

For the industrial production index the penalization is estimated (almost) infinitely high, which results in the linear shape of the trend functions. Figure 1 shows that there are differences between the trend estimations, although they are rather small. While the trend is linear for all residual autocorrelation structures there are differences in the intercept and the slope, especially when the results of the AR(1) and ARMA(1,1) error terms are compared to those of the AR(4) and ARMA(4,4) error terms. In detail the slope for the AR(1) case is 0.975 per year, whereas it is 1.025 in the case of an ARMA(4,4) error term.

To check for the robustness of the assumed autocorrelation structure Figure 2 displays the empirical autocorrelation- and partial autocorrelation functions for the residuals of the AR(1) and ARMA(4,4) estimation:

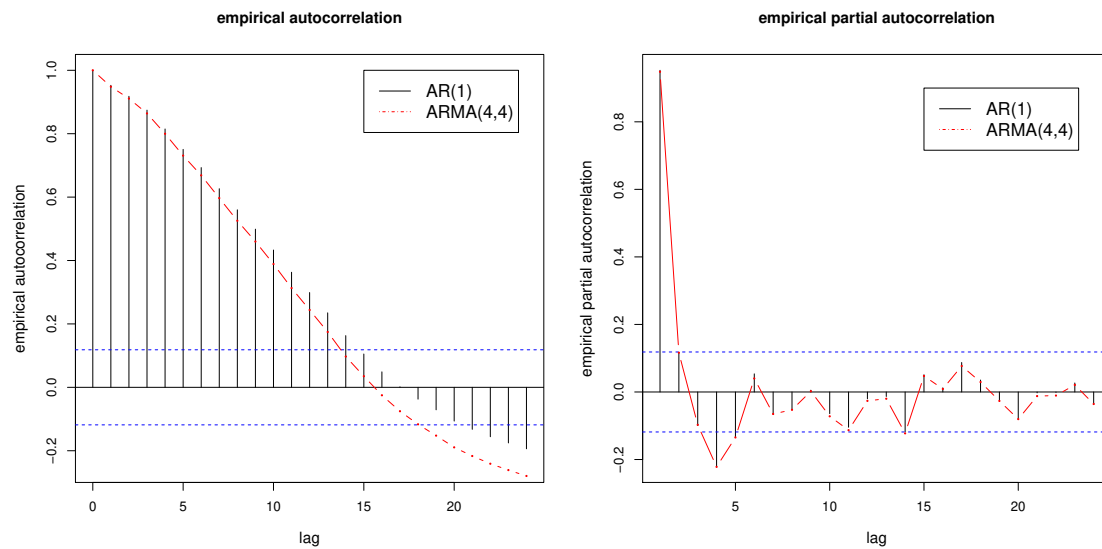


Fig. 2: Empirical autocorrelation and partial autocorrelation functions

The empirical autocorrelation and partial autocorrelation functions are almost equal for both cases. The empirical autocorrelation is damped and converges slowly to zero. The empirical partial autocorrelation adopts higher and clearly significant values for the first and fourth lag. Thus, higher autocorrelation structures rather than an AR(1)-process might be adequate, so that the results of the AR(4) or ARMA(4,4) cases appear to be preferable.

The analysis of the trend of the German industrial production by the HP-filter within a mixed model framework suggests that there is a very smooth and constant development on the long run. The trend estimation of the real industrial production as a measure for the economic development yields for every assumed residual autocorrelation structure a linear trend. These results are quite similar to those of Flaig (2005), who uses unobserved components models to examine the long run development of the German industrial production index, where the trend from 1991 to 2005 also shows a very smooth shape.

5 Conclusion

This paper reviews basic theory about the Hodrick-Prescott filter, penalized splines with a truncated polynomial basis as well as their link to linear mixed models. This link is advantageous, as it can be employed to derive an estimation for the penalization parameter.

Afterwards it is focused on the link between the HP-filter, penalized tp-splines and the Wiener-Kolmogorov filter. As the HP-filter is equal to a penalized tp-spline under certain settings of the parameters it can be incorporated into a mixed model framework. It is shown that for an autocorrelated error term the resulting model is equivalent to the Wiener-Kolmogorov filter where the trend is a second fold integrated random walk and the cycle follows a stationary autocorrelated process.

The possibility to incorporate the Hodrick-Prescott filter into a mixed model framework is used to estimate the trend component of the real German industrial production index. Although studies indicate that the exact assumption about the autocorrelation structure of the error term is of subordinate importance as long as the assumed structure doesn't deviate too strongly from the true one, the estimation is done for different residual autocorrelation structures. It turns out that the results are hardly affected by the assumed autocorrelation. In every case a very high penalization is estimated inducing a linear trend. Moreover, the empirical residual autocorrelation- and partial autocorrelation functions indicate that there is unlikely a massive misspecification in the assumed autocorrelation structure of the error term.

As a result this paper employs the link between penalized splines and the Hodrick-Prescott filter to proof that under certain settings of the model parameters penalized splines are equal to the Wiener-Kolmogorov filter with an autocorrelated cyclical component. Consequently, there is a direct link between penalized splines within the mixed model framework and unobserved components models with an integrated trend and an autocorrelated cyclical component.

Appendix

A Derivation of formula (8)

First, recall that for two invertible matrices \mathbf{A} and \mathbf{B} (e.g. Hamilton 1994 p.728)

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

This is used to derive

$$\begin{aligned} \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K})^{-1}\mathbf{Z}'\mathbf{y} &= [(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K})\mathbf{Z}^{-1}]^{-1}\mathbf{Z}'\mathbf{y} = \\ &= (\mathbf{Z}'\mathbf{Z}\mathbf{Z}^{-1} + \lambda\mathbf{K}\mathbf{Z}^{-1})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}' + \lambda\mathbf{K}\mathbf{Z}^{-1})^{-1}\mathbf{Z}'\mathbf{y} = \\ &= [\mathbf{Z}'^{-1}(\mathbf{Z}' + \lambda\mathbf{K}\mathbf{Z}^{-1})]^{-1}\mathbf{y} = (\mathbf{Z}'^{-1}\mathbf{Z}' + \lambda\mathbf{Z}'^{-1}\mathbf{K}\mathbf{Z}^{-1})^{-1}\mathbf{y} = \\ &= (\mathbf{I} + \lambda\mathbf{Z}'^{-1}\mathbf{K}\mathbf{Z}^{-1})^{-1}\mathbf{y}. \end{aligned}$$

B Proof that $\mathbf{Z}'^{-1}\mathbf{K}\mathbf{Z}^{-1} = \Delta_{\mu}'\Delta_{\mu}$

Given the special form of \mathbf{Z} for $l = 1$ and T equidistant knots, the inverse of \mathbf{Z} is in general given to (see also Paige 2010 p.870):

$$\mathbf{Z}^{-1} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Then the product of the transpose of \mathbf{Z}^{-1} and \mathbf{K} yields:

$$(\mathbf{Z}^{-1})'\mathbf{K} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix},$$

where the product of the two matrices has the dimension $T \times T$. This product can be decomposed in two matrices $\mathbf{E}_1 = \mathbf{0} \in \mathbb{R}^{T \times 2}$ and $\mathbf{E}_2 \in \mathbb{R}^{T \times T-2}$ so that $(\mathbf{Z}^{-1})' \mathbf{K} = (\mathbf{E}_1, \mathbf{E}_2)$. Having a closer look on \mathbf{E}_2 it becomes clear that $\mathbf{E}_2 = \Delta_\mu'$. Consequently, the product $(\mathbf{Z}^{-1})' \mathbf{K} = (\mathbf{0}, \Delta_\mu')$. Moreover, \mathbf{Z}^{-1} is partitioned into $(\mathbf{Z}_1', \mathbf{Z}_2')'$, with $\mathbf{Z}_1 \in \mathbb{R}^{2 \times T}$ and $\mathbf{Z}_2 \in \mathbb{R}^{T-2 \times T}$. Taking into consideration that $\mathbf{Z}_2 = \Delta_\mu$ it can finally be written:

$$(\mathbf{Z}^{-1})' \mathbf{K} \mathbf{Z}^{-1} = \begin{pmatrix} \mathbf{0}, \Delta_\mu' \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \Delta_\mu \end{pmatrix} = \mathbf{0} \mathbf{Z}_1 + \Delta_\mu' \Delta_\mu = \Delta_\mu' \Delta_\mu$$

References

- Bell, W.** (1984): *"Signal extraction for nonstationary time series,"* Annals of Statistics, 12, 646-664.
- Brumback, B. A., Ruppert, D., Wand, M. P.** (1999): Comment on *"Variable selection and function estimation in additive nonparametric regression using a data-based prior,"* Journal of the American Statistical Association, 94, 794-797.
- Claeskens, G., Krivobokova, T., Opsomer, J.** (2009): *"Asymptotic properties of penalized spline estimators,"* Biometrika, 96, 529-544.
- Danthine, J., Girardin, M.** (1989); *"Business Cycles in Switzerland. A Comparative Study,"* European Economic Review, 33(1), 31-50.
- Eilers, P., Marx, B.** (1996): *"Flexible smoothing with B-splines and penalties,"* Statistical Science, 11, 89-121.
- Fahrmeir, L., Kneib, T., Lang, S.** (2009): *"Regression - Modelle, Methoden und Anwendungen,"* Springer Verlag, Berlin.
- Flaig, G.** (2005): *"Time Series Properties of the German Production Index,"* Allgemeines Statistisches Archiv, 89, 419-434.
- Flaig, G.** (2012): *"Why we should use high values for the smoothing parameter of the Hodrick-Prescott filter,"* CESifo Working Paper Series No. 3816.
- Hamilton, J. D.** (1994): *"Time Series Analysis,"* Princeton University Press, Princeton/New Jersey.
- Harvey, A. C.** (1989): *"Forecasting, Structural Time Series Models and the Kalman Filter,"* Cambridge University Press, Cambridge.
- Hastie, T., Tibshirani, R.** (1990): *"Generalized Additive Models,"* Chapman and Hall, London.
- Hayes, K., Haslett, J.** (1999): *"Simplifying general least squares,"* American Statistician, 53, 376-381.
- Henderson, R.** (1924): *"On a new method of graduation,"* Transactions of the Actuarial Society of America, 25, 29-40.
- Hodrick, R. J., Prescott, E. C.** (1997); *"Post-War U.S. Business Cycles: An Empirical Investigation,"* Journal of Money, Credit and Banking, 29, 1-16.
- Kaiser, R., Maravall, A.** (2001): *"Measuring Business Cycles in Economic Time Series,"* Springer Verlag, New York.
- Kauermann, G., Krivobokova, T., Semmler, W.** (2011): *"Filtering Time Series with Penalized Splines,"* Studies in Nonlinear Dynamics & Econometrics, 15(2), Article 2.

- Kauermann, G., Opsomer, J.** (2011): "*Data-driven selection of the spline dimension in penalized spline regression,*" *Biometrika*, 98(1), 225-230.
- Kohn, R., Ansley, C., Wong, C.** (1992): "*Nonparametric spline regression with autoregressive moving average errors,*" *Biometrika*, 79, 335-346.
- Krivobokova, T., Kauermann, G.** (2007): "*A note on penalized spline smoothing with correlated errors,*" *Journal of the American Statistical Association*, 102, 1328-1337.
- Leser, C. E. V.**, (1961): "*A simple method of trend construction,*" *Journal of the Royal Statistical Society. Series B (Methodological)*, 23, 91-107.
- McCulloch, C. E., Searle, S. R.** (2001): "*Generalized, Linear, and Mixed Models,*" Wiley, New York.
- McElroy, T.** (2008): "*Matrix Formulas for Nonstationary ARIMA Signal Extraction,*" *Econometric Theory*, 24, 988-1009.
- O'Sullivan, F.** (1986): "*A statistical perspective on ill-posed inverse problems (c/r:P519-527),*" *Statistical Science*, 1, 502-518.
- Opsomer, J., Wang, Y., Yang, Y.** (2001): "*Nonparametric regression with correlated errors,*" *Statistical Science*, 16, 134-153.
- Paige, R. L.** (2010): "*The Hodrick-Prescott filter: A special case of penalized spline smoothing,*" *Electronic Journal of Statistics*, 4, 856-874.
- Pinheiro, J. C., Bates, D. M.** (2000): "*Mixed-Effects Models in S and S-PLUS,*" Springer-Verlag, New York.
- Proietti, T.** (2005): "*Forecasting and signal extraction with misspecified models,*" *Journal of Forecasting*, 24, 539-556.
- Proietti, T.** (2007): "*Signal Extraction and Filtering by Linear Semiparametric Methods,*" *Computational Statistics and Data Analysis*, 52, 935-958.
- Proietti, T., Luati, A.** (2007): "*Least Squares Regression: Graduation and Filters,*" Boumans, M. (Ed.), *Measurement in Economics: A Handbook*, ch. 16, Academic Press.
- Robinson, G. K.** (1991): "*That BLUP is a good thing: The estimation of random effects,*" *Statistical Science*, 6, 15-51.
- Ruppert, D.** (2002): "*Selecting the number of knots for penalized splines,*" *Journal of Computational and Graphical Statistics*, 11, 735-757.
- Ruppert, R., Wand, M., Carroll, R.** (2003): "*Semiparametric Regression,*" Cambridge University Press, New York.
- Schlicht, E.** (2005): "*Estimating the Smoothing Parameter in the so-called Hodrick-Prescott Filter,*" *Journal of the Japanese Statistical Society*, Vol. 35 No. 1, 99-119.
- Schlittgen, R., Streitenberg, B.** (2001): "*Zeitreihenanalyse,*" Oldenburg Wissenschaftsverlag, München, Wien.

Searle, S. R., Casella, G., McCulloch, C. E. (1992): *"Variance Components,"* Wiley, New York.

Vonesh, E. F., Chinchilli, V. M. (1997): *"Linear and Nonlinear Models for the Analysis of Repeated Measures,"* Marcel Dekker, New York.

Wang, Y. (1998): *"Mixed effects smoothing spline analysis of variance,"* Journal of the Royal Statistical Society, Series B, 60, 159-174

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

Whittle, P. (1983): *"Prediction and Regulation by Linear Least-Square Methods,"* University of Minnesota Press, Minneapolis.