THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 25_11

Wolfgang Polasek
Institute for Advanced Studies, Austria
University of Porto, Portugal
The Rimini Centre for Economic Analysis (RCEA), Italy

MCMC ESTIMATION OF EXTENDED HODRICK-PRESCOTT (HP) FILTERING MODELS

MCMC estimation of extended Hodrick-Prescott (HP) filtering models

Wolfgang Polasek[a]

[a] *Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria, and University of Porto, Rua Campo Alegre, Portugal*
[b] *Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria*

Abstract

The Hodrick-Prescott (HP) method was originally developed to smooth time series, i.e. to get a smooth (long-term) component. We show that the HP smoother can be viewed as a Bayesian linear model with a strong prior for the smoothness component. Extending this Bayesian approach in a linear model set-up is possible by a conjugate and a non-conjugate model using MCMC. The Bayesian HP smoothing model is also extended to a spatial smoothing model. We have to define spatial neighbors for each observation and we can use in a similar way a smoothness prior as for the HP filter in time series. The new smoothing approaches are applied to the (textbook) airline passenger data for time series and to the problem of smoothing spatial regional data. This new approach can be used for a new class of model-based smoothers for time series and spatial models.

Keywords: Hodrick-Prescott (HP) smoothers, Spatial econometrics, MCMC estimation, Airline passenger time series, Spatial smoothing of regional data, NUTS: nomenclature of territorial units for statistics.
JEL classification: C11, C15, C52, E17, R12 .

Email address: polasek@ihs.ac.at (Wolfgang Polasek)

1. Introduction

Regional data smoothing from a spatial point of view is an important issue for many applied regional scientists. In this paper, I consider the HP model from a Bayesian point of view and I show that the HP smoother is the posterior mean of a (conjugate) Bayesian linear regression model that uses a strong prior weight for the smoothness prior. The classicl approach to HP smoothing is reviewed in section 2 and the Bayesian version is introduced in section 3. I extend this model by introducing covariates in a larger regression model to define the extended HP (eHP) smoother in section 4, a model-based approach for data smoothers. Furthermore, I show that this approach allows define also a spatial smoothness concept that allows us to apply the Bayesian version of the HP filter to cross-sectional or regional data in section 5 and the spatial extended model is discussed in section 6. Both approaches are based on a distance concept to define spatial nearest neighbors (NN). An example for time series and for spatial regional GNP data in Europe will demonstrate this new smoothing approach in section 7. A final section concludes.

1.1. The HP filter for smoothing time series

The classical HP filter is a parametric estimation method to obtain a smooth trend component via the solution to the minimization of a loss function for a fixed (known) λ penalty parameter. There are 2 terms in the loss function. The first term in the loss function is a well-known measure of the goodness-of-fit, the error sum of squares (ESS). The second term punishes variations in the long-term trend component. The parameter λ is the key to the smoothing problem since it determines the trade-off between goodness-of-fit and the smoothness of the trend component. In the limit as $\lambda \to \infty$ the trend becomes as smooth as possible and eventually creates a sequence of parameter estimates that can be interpreted as cyclical component. When $\lambda \to 0$ then the trend component becomes equal to the data series y_t and the cyclical component approaches zero.

Many researchers have used the Hodrick and Prescott (1980, 1997) smoothing method (often called the HP filter). Hodrick and Prescott originally applied this procedure to post-war US quarterly data and their findings have since been extended in a number of papers including Kydland and Prescott (1990) and Cooley and Prescott (1995). Also the HP-filter is popular to analyse the business cycles and many researchers compare their results with those obtained for the US data. Blackburn and Ravn (1992) investigate UK business cycles, Danthine and Girardin (1989) the Swiss cycles, Dolado, Sebastian and Valles (1993), Puch and Licandro (1997) and Borondo, Gonzalez and Rodriguez (1999) study Spanish economic data, and Kim, Buckle and Hall (1994) look at data from New Zealand.

Hodrick and Prescott take λ as a fixed parameter, which they set equal to 1600 for US quarterly data. Their choice of this value was based upon a prior about the variability of the cyclical part relative to the variability of the change in the trend component. Hodrick and Prescott (1997, p.4) state that:

"If the cyclical components and the second differences of the growth components were identically and independently distributed, normal variables with means zero and variances σ_1^2 and σ_2^2 (which they are not), the conditional expectation of the τ, given the observations, would be the solution to [the minimisation problem (3)] when $\sqrt{\lambda} = \sigma_1/\sigma_2$. ... Our prior view is that a 5 percent cyclical component is moderately large, as is a one-eight of 1 percent change in the growth rate in a quarter. This led us to select $\sqrt{\lambda} = 5/(1/8)$ or $\lambda = 1600$."

Kydland and Prescott (1990, p. 9) argue further in favor of the choice of $\lambda = 1600$ for quarterly post war US data because:

"With this value, the implied trend path for the logarithm of real GNP is close to the one that students of the business cycle and growth would draw through a time plot of the series."

2. The HP filter as minimizer of a loss function

This section describes the HP smoothing problem from a classical point of view of parameter estimation. Starting point is the following (overparameterized) regression problem for the observations $\mathbf{y} = [y_1, ..., y_T]'$

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon} \quad with \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T], \tag{1}$$

where the HP smoother is defined as parameter vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_T]^{\top}$. The classical approach for this problem is based on an optimisation of a special loss function:

Definition 1 (The smoothed squared loss (SSL) function). *To obtain a HP-type smoother for the observations* $\mathbf{y}$ *in model (1) we define the smoothed squared loss (SSL) function that yields the smoother* $\hat{y}$*:*

$$\hat{\mathbf{y}} = \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad with \quad SSL(\boldsymbol{\tau}) = ESS(\boldsymbol{\tau}) \ + \ \lambda * smooth(\boldsymbol{\tau}) \tag{2}$$

where the ESS is defined as error sum of squares of the ideo-parameterized (i.e. equal sized) and homoskedastic regression model:

$$ESS(\boldsymbol{\tau}) = \sum_t (y_t - \boldsymbol{\tau}_t)^2.$$

The $smooth(\boldsymbol{\tau})$ *is a (quadratic) penalty function on the roughness of the fit:* $smooth(\boldsymbol{\tau}) = [\Delta_k(\boldsymbol{\tau})]^2$*, where* $\Delta_k(\boldsymbol{\tau})$ *can be a differencing function of fixed order (usually* $k = 2$*) between neighboring observations of* $\mathbf{y}$*. (Note that the notion of neighbors assumes a metric for all the observations in* $\mathbf{y}$*.)* λ *is the known penalty parameter for the smooth.*

The original HP filter problem can be defined as a minimizer of the smoothed square loss (SSL) function, which has two components, the goodness of fit and the smooth: $SSL = ESS \ + \ \lambda * smooth$ or

$$\hat{\boldsymbol{\tau}} = \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad with \quad SSL(\boldsymbol{\tau}) = \sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=1}^{T} (\Delta^2 \tau_t)^2. \tag{3}$$

The solution to this problem is given by the next theorem.

Theorem 1. *[The HP smoother as posterior mean]*

We consider the regression problem in (1) and we like to obtain the minimum SSL estimate of $\boldsymbol{\tau}$ under the SSL function as in Definition 1. The minimum of the SSL function is under the assumption of a normal distribution given by

$$\min_{\boldsymbol{\tau}} [(\mathbf{y}-\boldsymbol{\tau})^\top(\mathbf{y}-\boldsymbol{\tau}) + \lambda\boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau}] = \boldsymbol{\tau}_{**}, \tag{4}$$

which is the posterior mean[1] *of the equivalent Bayesian model*

$$\boldsymbol{\tau}_{**} = [\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K}]^{-1}\mathbf{y} = \mathbf{A}_{**}\mathbf{y}. \tag{5}$$

with the posterior precision matrix

$$\mathbf{A}_{**}^{-1} = \mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K}. \tag{6}$$

The second order[2] *differencing matrix $\mathbf{K} : (T-2)\times T$ is given by*

$$\mathbf{K} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & ... & 0 & 0 & 0 \\ & ... & & ... & ... & ... & ... & ... & \\ 0 & 0 & 0 & 0 & 0 & ... & 1 & -2 & 1 \end{pmatrix} \tag{7}$$

Proof 1. *The proof relies on rewriting the SSL function $SSL = ESS + \lambda * smooth$ as a sum of 2 quadratic forms in $\boldsymbol{\tau}$:*

$$ESS(\boldsymbol{\tau}) = (\mathbf{y}-\boldsymbol{\tau})^\top(\mathbf{y}-\boldsymbol{\tau}) \quad \text{and} \quad smooth(\boldsymbol{\tau}) = \boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau} \tag{8}$$

and to apply Theorem 9 of the appendix:

$$(\mathbf{y}-\boldsymbol{\tau})^\top(\mathbf{y}-\boldsymbol{\tau}) + \lambda\boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau} = (\boldsymbol{\tau}-\boldsymbol{\tau}_{**})'(\boldsymbol{\tau}-\boldsymbol{\tau}_{**}) + \mathbf{y}^\top\lambda\mathbf{K}^\top\mathbf{K}(\lambda\mathbf{K}^\top\mathbf{K}+\mathbf{I}_T)^{-1}\mathbf{I}_T\mathbf{y} \tag{9}$$

where $\mathbf{I}_T$ is a $T\times T$ identity matrix, and $\mathbf{K} = \{k_{ij}\}$ is a $(T-2)\times T$ matrix with elements given by

$$k_{ij} = \begin{cases} 1 & \text{if} \quad i=j \quad \text{or} \quad j=i+2, \\ -2 & \text{if} \quad j=i+1, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The second quadratic form is centered around zero, therefore the posterior mean $\boldsymbol{\tau}_{**}$ has a simple form in (5). From the combination of quadratic forms we see that only the first term involves $\boldsymbol{\tau}$, while the second is independent of $\boldsymbol{\tau}$. Therefore the whole expression is minimized if the first term is set to zero and $\boldsymbol{\tau}$ is set equal to the posterior mean $\boldsymbol{\tau}_{**}$. For the HP smoother the equivalent Bayesian model is therefore the following informative normal (homoskedastic) regression model:

$$\mathbf{y} \sim \mathcal{N}[\boldsymbol{\tau}, \sigma^2\mathbf{I}_T] \quad \text{with} \quad \mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}]. \tag{11}$$

2.1. Properties of the HP smoothness filter

The inversion of the posterior precision matrix $\mathbf{A}_{**}^{-1} = \mathbf{I}_T + \mathbf{K}'\lambda\mathbf{K}$ follows the inversion lemma[3]

[3] $(A+BCB')^{-1} = A^{-1} - A^{-1}B(C^{-1}+B'A^{-1}B)^{-1}B'A^{-1}$

$$\mathbf{A}_{**} = (\mathbf{I}_T + \mathbf{K}'\lambda\mathbf{K})^{-1} = \mathbf{I}_T - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}. \tag{12}$$

For the HP smooth in (5) we find now

$$\begin{aligned} \mathbf{y}_{**} &= (\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K})^{-1}\mathbf{y} = [\mathbf{I}_T - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}]\mathbf{y} \\ &= \mathbf{y} - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}\mathbf{y} = \mathbf{y} - \hat{\mathbf{e}}. \end{aligned} \tag{13}$$

The second term $\hat{\mathbf{e}} = \mathbf{P}_\lambda\mathbf{y}$ with the projector

$$\mathbf{P}_\lambda = \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K} \tag{14}$$

estimates the rough or noise component of this smoothness problem:

$$data \;= fit \;+\; rough \quad or \quad \mathbf{y} = \mathbf{y}_{**} + \hat{\mathbf{e}}.$$

A simple measure for the size of the smoothing is the variance of the rough: $Var(\hat{\mathbf{e}}) = \sum_t \hat{e}_t^2/T$. Note that the mean of $\hat{\mathbf{e}}$ is zero since $\mathbf{K}\mathbf{1}_T = \mathbf{1}_T'\mathbf{K}' = \mathbf{0}$ and therefore we have the property $\bar{\mathbf{y}} = \bar{\mathbf{y}}_{**}$, which is also found for least squares (LS) decompositions.

3. The HP filter as Bayesian smoothness model

In the Bayesian framework, we also start from the regression model (15)

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{I}_T], \tag{15}$$

with the identity matrix as "regressors" and where $\boldsymbol{\tau} : T \times 1$ is the equal-sized parameter (or ideo-parameter) vector to be estimated and the error term $\boldsymbol{\varepsilon}$ is assumed to be homoskedastic. The prior is obtained in the following way: we specify for $\boldsymbol{\tau}$ a prior density for a transformed parameter model, where the transformation is the second order differencing matrix $\mathbf{K} : (T-2) \times T$:

$$\mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}]. \tag{16}$$

In this special case with prior mean $\mathbf{0}$ it is easy to see that the prior is equivalent to[4] the distributional assumption

$$\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)(\mathbf{K}^\top\mathbf{K})^{-1} = \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{A}_*] \quad with \quad \mathbf{A}_* = (\lambda\mathbf{K}^\top\mathbf{K})^{-1}. \tag{17}$$

The problem with the distribution in (17) is that the covariance matrix $\mathbf{A}_* = (\lambda\mathbf{K}^\top\mathbf{K})^{-1}$ is not of full rank and defines a singular, rank deficient normal distribution[5]. But this problem of rank deficiency of the prior is not a problem in a Bayesian analysis, as long as the likelihood function is normally distributed with full rank covariance matrix: the posterior precision is the sum of 2 precision matrices where at least one of them must have full rank.

Since λ is in the denominator it has the form of an hypothetical sample size $n' = \lambda$. In a typical regression application we give the prior information only a small weight, like the equivalent of 1 or 2 sample points. Thus, if we specify a large λ, then this means that we give the prior density a much larger weight than the sample mean (or likelihood). In this case the posterior mean (or HP) smooth is shifted to the prior location, which is zero, but in the transformed (= differenced) form of the model. This means that the fit is smoothed towards a function that minimizes the second order difference of the $\boldsymbol{\tau}$.

It is interesting to note that classical and Bayesian smoothing requires strong prior information. In Bayesian terms this is made explicit while in classical terms this information is implicitly hidden in the term "smoothing parameter". But strong priors follows the opposite principle of objectivity or non-involvement that is so often promoted in the case of inference for regression coefficients: For inference we try to minimize the influence of the prior (small n'), while for the smoothing problem we maximize the influence of the prior (large $n' = \lambda$).

Following the textbook Bayesian regression approach, the posterior mean of the parameters $\boldsymbol{\mu}$ is given by the usual combination of prior and likelihood and relies on the algebraic solution of Theorem 9.

This is a matrix weighted average between the prior location $\mathbf{0}$ and the ML location y. Note that in the Bayesian framework it does not matter that the τ parameter has T components, as many as there are observations, as long as there is a proper prior distribution.

3.1. Conjugate multi-normal-gamma (mNG) inference for HP smoothing

First, we describe the conjugate smoothing approach that is in analogy to the Normal-Wishart sampling (NWS) model that can be found in Polasek (2010) and is listed the appendix.

[4] $p(\boldsymbol{\tau}) \propto exp[-0.5(\mathbf{K}\boldsymbol{\tau})^\top(\mathbf{K}\boldsymbol{\tau})\lambda/\sigma^2] = exp[-0.5\boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau}\lambda/\sigma^2] \propto \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)(\mathbf{K}^\top\mathbf{K})^{-1}]$

[5] Note that the inverse does formally not exist and therefore it is more elegant to define the multivariate normal distribution for such cases by the precision matrix.

We consider the conjugate normal-gamma model for the inference of an unknown mean $\boldsymbol{\mu}$ in a univariate sampling problem (with sample size n):

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T], \quad or \quad \mathbf{y} \sim \mathcal{N}[\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma}_0]. \tag{18}$$

To emphasize the similarity of the HP smoothing model with the Bayesian model where the prior is assigned a hypothetical sample size, we set $\lambda = n'$ in the following theorem.

Theorem 2. *[The multivariate normal-gamma sampling (mNGS) model]*
We consider the smoothing model in (18) with prior density as in (17), then the conjugate Bayesian inference can be done in the following way.

The prior distribution is given as a normal-gamma density

$$(\boldsymbol{\mu}, \sigma^{-2}) \sim \mathcal{N}_n\Gamma[\boldsymbol{\mu}_*, \mathbf{A}_*, s_*^2, n_*]$$

and the likelihood of the observed data in the set

$$\mathcal{Y} = \{\mathbf{y}_i \sim \mathcal{N}[\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma}_0], \quad i = 1, \ldots, n\}$$

yields the posterior distribution

$$(\boldsymbol{\mu}, \sigma^{-2}) \mid \mathbf{x} \sim \mathcal{N}_n\Gamma[\boldsymbol{\mu}_{**}, \mathbf{A}_{**}, s_{**}^2, n_{**}].$$

with the parameters

$$\begin{aligned}
\boldsymbol{\mu}_{**} &= \mathbf{A}_{**}(n'\mathbf{K}^\top\mathbf{K}\boldsymbol{\mu}_* + \boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{y}}), \\
\mathbf{A}_{**}^{-1} &= n'\mathbf{K}^\top\mathbf{K} + \boldsymbol{\Sigma}_0^{-1}, \\
n_{**} &= n_* + n, \\
\alpha &= \mathbf{y}^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K} + \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\Sigma}_0\mathbf{y} \\
n_{**}s_{**}^2 &= \quad n_* s_*^2 + ns^2 + \alpha
\end{aligned}$$

The current error sum of squares is $ns^2 = (\mathbf{y} - \boldsymbol{\mu})^\top(\mathbf{y} - \boldsymbol{\mu})$ and α is the discrepancy term that serves as a penalty term for the variance in all conjugate models.

Proof 2.
The likelihood of the above smoothing model (18) is simply derived from $\mathbf{y} \sim \mathcal{N}[\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}_0]$.
Next we define a special 'multi-normal-gamma' or family of mNG conjugate distribution that follows from the normal-gamma ($\mathcal{N}\Gamma$) distribution.

$$(\boldsymbol{\mu}, \sigma^{-2}) \sim \mathcal{N}_n\Gamma[\boldsymbol{\mu}_*, \mathbf{A}_*, \sigma_*^2\boldsymbol{\Sigma}_0, n_*], \tag{19}$$

where $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$ is a known covariance matrix. (A normal-Wishart (NW) distribution can also be assumed but the posterior information for the covariance matrix is very weak because there is only one observation.)

Similar as for the mNΓ distribution we define the mNΓ distribution as

$$\begin{aligned} p(\boldsymbol{\mu}, \sigma^{-2}) &= p(\boldsymbol{\mu} \mid \sigma^{-2}) p(\sigma^{-2}) = \mathcal{N}[\boldsymbol{\mu} \mid \boldsymbol{\mu}_*, \sigma^2/n'(\mathbf{K}^\top \mathbf{K})^{-1}] \; \Gamma[\sigma^{-2} \mid s_*^2, n_*] \\ &\propto exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_*)' n' \mathbf{K}^\top \mathbf{K}(\boldsymbol{\mu}-\boldsymbol{\mu}_*)\right)\right\} exp\left\{-\frac{1}{2\sigma^2} n_* s_*^2\right\}. \end{aligned} \tag{20}$$

Therefore the pdf of the mNΓ $= \mathcal{N}_n\Gamma$ distribution has the following form

$$p(\boldsymbol{\mu}, \sigma^{-2}) \propto (\sigma^{-2})^{\frac{n+n_*}{2}-1} exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_*)' n' \mathbf{K}^\top \mathbf{K}(\boldsymbol{\mu}-\boldsymbol{\mu}_*) + n_* s_*^2\right)\right\}.$$

This has the structure of a NΓ distribution[6] *but only the $\boldsymbol{\mu}$ vector is n-dimensional. Now we find the posterior mNΓ distribution by multiplying the prior with the likelihood:*

$$\begin{aligned} p(\boldsymbol{\mu}, \sigma^{-2} \mid \mathbf{X}) &\propto (\sigma^{-2})^{\frac{n+n_*}{2}-1} exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_*)' n' \mathbf{K}^\top \mathbf{K}(\boldsymbol{\mu}-\boldsymbol{\mu}_*) + n_* s_*^2\right)\right\} \\ &\cdot \; exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \\ &\propto \mathcal{N}_n\Gamma[\boldsymbol{\mu}_{**}, \mathbf{A}_{**}, \sigma_{**}^2 \boldsymbol{\Sigma}_0, n_{**}]. \end{aligned} \tag{22}$$

We have to apply the theorem of combining the 2 quadratic forms in $\boldsymbol{\mu}$ (see Appendix) to get

$$(\boldsymbol{\mu}-\boldsymbol{\mu}_{**})^\top \mathbf{H}_{**}(\boldsymbol{\mu}-\boldsymbol{\mu}_{**}) + (\mathbf{y}-\boldsymbol{\mu}_*)^\top n' \mathbf{K}^\top \mathbf{K}(n' \mathbf{K}^\top \mathbf{K} + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}(\mathbf{y}-\boldsymbol{\mu}_*) \tag{23}$$

The second term is called discrepancy term between the observation $\mathbf{y}$ and the prior location which is zero. Thus the discrepancy is for $\boldsymbol{\mu}_* = \mathbf{0}$ given by

$$\alpha = \mathbf{y}^\top n' \mathbf{K}^\top \mathbf{K}(n' \mathbf{K}^\top \mathbf{K} + \boldsymbol{\Sigma})_0^{-1} \boldsymbol{\Sigma}_0 \mathbf{y},$$

and the parameters $\boldsymbol{\mu}_{**}$ and $\mathbf{A}_{**}$ are given as in (19).

The posterior multi-normal-gamma $\mathcal{N}_n\Gamma$ density can be factored as

$$p(\boldsymbol{\mu} \mid \sigma^{-2}) = \mathcal{N}[\boldsymbol{\mu} \mid \boldsymbol{\mu}_{**}, \sigma_{**}^2 = \sigma^2/n''] \; \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$$

with the marginal distribution for $\boldsymbol{\mu}$ being a t-distribution with n_{**} d.f. given by

$$\boldsymbol{\mu} \mid \mathbf{y} \sim t\left[\boldsymbol{\mu}_{**}, s_\mu^2 = \frac{s_{**}^2}{n''}, n_{**}\right] \tag{24}$$

The smoothness predictor of the $\mathbf{y}$ observations is in the Bayesian case is given by the posterior distribution of $\boldsymbol{\mu}$. The point estimate of the smoother is the point estimate of the posterior distribution. A common

[6] Recall that the NΓ $[\boldsymbol{\mu}_*, \mathbf{I}_n, s_*^2 n_*]$ prior distribution is defined as

$$p(\boldsymbol{\mu}, \sigma^{-2}) \propto (\sigma^{-2})^{\frac{n+n_*}{2}-1} exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_*)^\top(\boldsymbol{\mu}-\boldsymbol{\mu}_*) + n_* s_*^2\right)\right\} \tag{21}$$

choice is the posterior mean which is given by (19)

$$\boldsymbol{\mu}_{**} = \mathbf{A}_{**}(n'\mathbf{K}^\top\mathbf{K}\boldsymbol{\mu}_* + \boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{y}}), \tag{25}$$

For one observation $\mathbf{y}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$ this is the same formula as in the classical case in (13): $\hat{\mathbf{y}} = \boldsymbol{\mu}_{**}$

$$\boldsymbol{\mu}_{**} = (\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K})^{-1}\mathbf{y} = [\mathbf{I}_T + \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}]\mathbf{y}. \tag{26}$$

The reason is that we have only one observation for inference and that the smoothness assumption is brought into the classical model in the same way as Bayesian enter their prior information.

The smoothed series is obtained by prediction, where the point prediction is obtained again via the posterior mean as in (25).

3.2. MCMC: A non-conjugate Bayesian HP smoother

Now we show how MCMC can be used to produce a non-conjugate Bayesian HP smoother.

Theorem 3. *[MCMC for HP-smoothing for non-conjugate priors]*

The posterior simulator of the parameters $\theta = (\boldsymbol{\tau}, \sigma^{-2})$ for the simple HP smoothing model (15) with prior (17) is given by the following iteration:

1. *Get a starting value for $\sigma^2 = Var(\mathbf{y})$;*
2. *Draw $\boldsymbol{\tau}$ from $\mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \mathbf{A}_{**}]$;*
3. *Draw σ^{-2} from $\Gamma[\sigma^{-2} \mid s^2_{**}n_{**}/2, n_{**}/2]$;*
4. *Repeat until convergence.*

The hyper-parameters of the fcd's can be found in the proof: (27) and (28).

Proof 3. 1. *The fcd for the residual precision σ^{-2}*

$$p(\sigma^{-2} \mid \boldsymbol{\tau}, \mathbf{y}) \quad \propto \quad \Gamma\left[\sigma^{-2} \mid s^2_{**}, n_{**}\right]$$

we find a gamma distribution with the parameters

$$\begin{aligned} n_{**} &= n_* + n, \\ n_{**}s^2_{**} &= n_* s^2_* + \sum_{i=1}^{n}(y_i - \tau_i)^2 \end{aligned} \tag{27}$$

2. *The fcd for the $\boldsymbol{\tau}$ coefficients is*

$$\begin{aligned} p(\boldsymbol{\tau} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\tau} \mid \mathbf{0}, \mathbf{A}_*] \cdot \mathcal{N}[\mathbf{y} \mid \boldsymbol{\tau}, \sigma^2\mathbf{I}_T] \\ &= \mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \mathbf{A}_{**}] \end{aligned}$$

with the parameters $\boldsymbol{\tau}_{**} = [\mathbf{I}_T + \lambda \mathbf{K}^\top \mathbf{K}]^{-1} \mathbf{y} = \mathbf{A}_{**} \mathbf{y}$ *and*

$$\mathbf{A}_{**}^{-1} = \mathbf{A}_{*}^{-1} + \sigma^{-2} \mathbf{I}_n. \tag{28}$$

4. The extended regression and smoothing model (R'n'S: regression and smoothing)

In this section we extend the smoothing model in (1) to a more general regression framework, where the additional regressors control for other (ideosyncratic) influences:

$$\mathbf{y} = \mathbf{I}_T \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T]. \tag{29}$$

The conditional mean (or fit) of this model is now defined by

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{I}_T \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} = [\mathbf{I}_T : \mathbf{X}]\boldsymbol{\gamma} = \mathbf{Z}\boldsymbol{\gamma} \\ & \quad \textit{with} \quad \mathbf{Z} = [\mathbf{I}_T : \mathbf{X}] \quad \textit{and} \quad \boldsymbol{\gamma}^\top = (\boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top). \end{aligned} \tag{30}$$

Note that now we have $T + p$ parameters to estimate in $\boldsymbol{\gamma}$ since $\boldsymbol{\beta} : p \times 1$. The classical approach is based on an optimisation problem with second order smoothness restriction similar to the Definition 1

$$\min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad \textit{with} \quad SSL(\boldsymbol{\tau}) = \sum_{t=1}^{T} (y_t - \mu_t)^2 + \lambda \sum_{t=1}^{T} (\Delta^2 \mu_t)^2. \tag{31}$$

The penalty term uses the first and second differences of the μ parameter:

$$\Delta \mu_t = \mu_t - \mu_{t-1} = \tau_t - \tau_{t-1} + (x_t - x_{t-1})\beta, \quad for \quad t = 1, ..., T, \tag{32}$$

and $\Delta^2 \mu_t = \Delta \mu_t - \Delta \mu_{t-1}$.

4.1. The Bayesian extended HP smoothness model

In this section we discuss the extended HP smoothing problem (eHP) from a classical and a Bayesian point of view.

Definition 2 (The smoothed squared loss (SSL) function for extended regression). *We consider the extended (homoskedastic) regression model* $\mathbf{y} = \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *as in (29). Conditional on* $\boldsymbol{\beta}$*, the SSL function stays the same, only the ESS function changes and includes the regression term of the extended model:*

$$ESS(\boldsymbol{\tau} \mid \boldsymbol{\beta}) = \sum_i (y_i - \boldsymbol{\tau}_i - \mathbf{x}_i\boldsymbol{\beta})^2,$$

where $\mathbf{x}_i$ *is the i-th row of the regressor matrix* $\mathbf{X}$*. This yields the smoother* $\hat{\mathbf{y}}_\beta$*:*

$$\hat{\mathbf{y}}_\beta = \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau} \mid \boldsymbol{\beta}) \quad with \quad SSL(\boldsymbol{\tau} \mid \boldsymbol{\beta}) = ESS(\boldsymbol{\tau} \mid \boldsymbol{\beta}) \ + \ \lambda * smooth(\boldsymbol{\tau}) \tag{33}$$

where the smooth is the quadratic penalty function as in Definition 1.

From this definition we see that a joint minimum SLL estimate can be found by minimizing over the joint parameters $(\boldsymbol{\tau}, \boldsymbol{\beta})$. This is not the same as the HP smoother of the residuals when we purge (by regression) from the $\mathbf{y}$ the $\mathbf{X}\boldsymbol{\beta}$ component. Let the OLS residuals be $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ with $\mathbf{X}\boldsymbol{\beta}$ the OLS estimate, then $\hat{\mathbf{u}}_{HP}$ can be obtained from Definition 1. But $\hat{\mathbf{u}}_{HP} \neq \hat{\mathbf{y}}_{eHP}$ as can be seen from in the application in Figure 1. Therefore the eHP method allows to generalize the HP approach to models with trends, outliers or other types of breaks or regime shifts.

For the Bayesian solution we have to construct a prior distribution for $\boldsymbol{\gamma}$ that uses 2 hypothetical sample sizes, λ is the one for the $\boldsymbol{\tau}$, and n_2 for the regression parameters $\boldsymbol{\beta}$.

The Bayesian approach for the extended HP filtering problem with additional regressors is straight forward. Using the stacked $\boldsymbol{\gamma}$ parameter we apply conjugate normal-gamma model for the inference of an unknown mean μ in a univariate sampling problem:

Definition 3 (eHP: The Bayesian HP smoother for the extended regression model). *We consider the normal linear regression model*

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T], \quad or \quad \mathbf{y} \sim \mathcal{N}[\mathbf{Z}\boldsymbol{\gamma}, \sigma^2 \boldsymbol{\Sigma}_0], \tag{34}$$

with $\boldsymbol{\gamma} = \begin{pmatrix} \tau \\ \beta \end{pmatrix}$ *and where* $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$ *is a known covariance matrix.*
Now we use a special 'multi-NG' conjugate distribution that uses blocks from the $N\Gamma$ *distribution. The prior is*

$$(\boldsymbol{\gamma}, \sigma^{-2}) \sim \mathcal{N}_{n+p}\Gamma[\boldsymbol{\gamma}_*, \mathbf{A}_*, \sigma_*^2, n_*], \quad with \quad \mathbf{A}_* = diag(\lambda \mathbf{K}^\top \mathbf{K}, n_2 \mathbf{I}_p)^{-1} = \begin{pmatrix} (\lambda \mathbf{K}^\top \mathbf{K})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p / n_2 \end{pmatrix} \tag{35}$$

being a block-diagonal matrix. λ *is the large hypothetical sample size for the* τ *parameter, with the covariance matrix* $\mathbf{K}^\top\mathbf{K}$ *that derives from the second order smoothness assumption, and the small* n_2 *for the rather non-informative prior information for the* $\boldsymbol{\beta} : p \times 1$ *regression coefficients.*

The Bayesian inference with conjugate normal-gamma distributions is shown in the next theorem.

Theorem 4. *[The conjugate extended HP smoothing model*
We consider the extended HP smoothing model in (34) with parameters $\theta = (\boldsymbol{\gamma}, \sigma^{-2})$ *as in Definition 3. The conjugate Bayesian inference follows the following steps:*

The prior distribution is given as a multi-normal-gamma (mNG) density

$$(\boldsymbol{\gamma}, \sigma^{-2}) \sim \mathcal{N}_{n+p}\Gamma[\boldsymbol{\gamma}_*, \mathbf{A}_*, s_*^2, n_*]$$

and the likelihood of the data

$$\mathbf{y} \sim \mathcal{N}[\mathbf{Z}\boldsymbol{\gamma}, \sigma^2 \boldsymbol{\Sigma}_0]$$

yields the posterior distribution

$$(\boldsymbol{\gamma}, \sigma^{-2}) \mid \mathbf{y} \sim \mathcal{N}_n\Gamma[\boldsymbol{\gamma}_{**}, \mathbf{A}_{**}, s_{**}^2, n_{**}].$$

with the parameters

$$\begin{aligned} \boldsymbol{\gamma}_{**} &= \mathbf{A}_{**}(\mathbf{A}_*\boldsymbol{\gamma}_* + \boldsymbol{\Sigma}_0^{-1}\mathbf{y}), \\ \mathbf{A}_{**}^{-1} &= \mathbf{A}_*^{-1} + \boldsymbol{\Sigma}_0^{-1}, \\ n_{**} &= n_* + n, \\ n_{**}s_{**}^2 &= n_* s_*^2 + ns^2 + \alpha \\ \alpha &= \mathbf{y}^\top \mathbf{A}_*(\mathbf{A}_* + \boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0\mathbf{y} \end{aligned}$$

The current error sum of squares is $ns^2 = (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})^\top(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})$ and α is the discrepancy term that serves as a penalty term for the variance in all conjugate models.

Proof 4.

The likelihood of the above smoothing model (18) is simply derived from $\mathbf{y} \sim \mathcal{N}[\mathbf{Z}\boldsymbol{\gamma}, \sigma^2\boldsymbol{\Sigma}_0]$*.*
The joint prior for the eHP model is under block independence

$$\begin{aligned} p(\boldsymbol{\gamma}, \sigma^{-2}) &\propto (\sigma^{-2})^{\frac{n+n_*}{2}-1} exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\gamma}-\boldsymbol{\gamma}_*)'\lambda\mathbf{K}^\top\mathbf{K}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_*) + n_* s_*^2\right)\right\} \\ & \quad exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_*)' n_2 \mathbf{I}_p(\boldsymbol{\gamma}-\boldsymbol{\gamma}_*)\right\}. \end{aligned}$$

We find the posterior $m\mathcal{N}\Gamma$ *distribution by multiplying the prior with the likelihood:*

$$\begin{aligned} p(\boldsymbol{\gamma}, \sigma^{-2} \mid \mathcal{X}) &\propto (\sigma^{-2})^{\frac{n+n_*}{2}-1} exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\gamma}-\boldsymbol{\gamma}_*)'\lambda\mathbf{K}^\top\mathbf{K}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_*) + n_* s_*^2\right)\right\} \\ & \quad \cdot \quad exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\gamma})'\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}-\boldsymbol{\gamma})\right\} \\ &\propto \mathcal{N}_n\Gamma[\boldsymbol{\gamma}_{**}, \mathbf{A}_{**}, \sigma^2_{**}, n_{**}]. \end{aligned} \tag{36}$$

We have to apply Theorem 9 for combining quadratic forms (see Appendix)

$$\begin{aligned} & (\mathbf{y}-\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}-\boldsymbol{\gamma}) + (\boldsymbol{\gamma}-\boldsymbol{\gamma}_*)^\top\lambda\mathbf{K}^\top\mathbf{K}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_*) = \\ = \; & (\boldsymbol{\gamma}-\boldsymbol{\gamma}_{**})^\top\mathbf{H}_{**}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_{**}) + (\mathbf{y}-\boldsymbol{\gamma}_*)^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K}+\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0(\mathbf{y}-\boldsymbol{\gamma}_*). \end{aligned} \tag{37}$$

The second term is called discrepancy term between the $\boldsymbol{\gamma}$ *and the prior location which is zero. This discrepancy term is for* $\boldsymbol{\gamma}_* = \mathbf{0}$ *given by*

$$\alpha = \mathbf{y}^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K}+\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0\mathbf{y},$$

and the parameters $\boldsymbol{\mu}_{**}$ *and* $\tilde{\mathbf{A}}_{**}$ *are given as in (19). Note that the posterior precision can be written in a partitioned form*

$$\begin{aligned} \tilde{\mathbf{A}}_{**}^{-1} &= \mathbf{A}_*^{-1} + \mathbf{Z}'\boldsymbol{\Sigma}_0\mathbf{Z} = \\ &= \begin{pmatrix} \lambda\mathbf{K}^\top\mathbf{K} & 0 \\ 0 & n_2\mathbf{I}_p \end{pmatrix} + \begin{pmatrix} \mathbf{I}_n & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{X}^\top\mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n + \lambda\mathbf{K}^\top\mathbf{K} & \mathbf{X} \\ \mathbf{X}^\top & n_2\mathbf{I}_p + \mathbf{X}^\top\mathbf{X} \end{pmatrix}. \end{aligned} \tag{38}$$

The posterior mean in the eHP model is given by

$$\boldsymbol{\gamma}_{**} = \begin{pmatrix} \tau_{**} \\ \beta_{**} \end{pmatrix} = \tilde{\mathbf{A}}_{**}\begin{pmatrix} \mathbf{y} \\ \mathbf{X}^\top\mathbf{y} \end{pmatrix} \tag{39}$$

For simplification we briefly discuss the semi-informative smoothing model for $n_2 = 0$.

Theorem 5 (The semi-conjugate HP smoother for the extended regression model). *We consider the model (34) with demeaned (centered)* $\mathbf{y}$ *and* $\mathbf{X}$ *variables, so that* $\mathbf{KX} \neq \mathbf{0}$*. Furthermore we assume a prior as in (35) with* $n_2 = 0$*, which we will call 'partial informative' or 'semi-conjugate' HP model.*

The extended HP smoother in (64) is given by

$$\boldsymbol{\tau}_{**} = \mathbf{y}_{HP} - \mathbf{X}_{HP}\hat{\boldsymbol{\beta}}_\lambda, \quad \textit{with} \tag{40}$$

$$\hat{\boldsymbol{\beta}}_\lambda = (\mathbf{X}^\top \mathbf{P}_\lambda \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{P}_\lambda \mathbf{y}, \tag{41}$$

with $\mathbf{A}_{**} = \mathbf{I}_n - \lambda \mathbf{K}^\top \mathbf{K}$ *and* $\mathbf{P}_\lambda$ *in (14). The simple HP smoother of* $\mathbf{y}$ *and* $\mathbf{X}$ *are given by*

$$\mathbf{y}_{HP} = (\mathbf{I}_n + \lambda \mathbf{K}^\top \mathbf{K})^{-1}\mathbf{y}(= \widehat{\boldsymbol{\tau}}) \tag{42}$$

$$\mathbf{X}_{HP} = (\mathbf{I}_n + \lambda \mathbf{K}^\top \mathbf{K})^{-1}\mathbf{X}(= \hat{\mathbf{X}}). \tag{43}$$

The second term $\mathbf{X}_{HP}\hat{\mathbf{b}}_\lambda$ *acts as a correction term to the original HP smoothing problem (i.e.* $\mathbf{y}_{HP}$*) without the regression part* $\mathbf{X}\boldsymbol{\beta}$*. The correction term is a special prediction vector of a difference-purged regression model*

$$\mathbf{y} = \mathbf{X}_{\langle K \rangle}\boldsymbol{\beta}_\lambda + \mathbf{u} \quad \textit{with} \quad \mathbf{u} \sim \mathcal{N}[\mathbf{0}, \sigma_u^2 \mathbf{P}_\lambda^{-1}] \quad \textit{and} \quad \mathbf{X}_{\langle K \rangle} = \mathbf{P}_\lambda^{1/2}\mathbf{X}, \tag{44}$$

where $\mathbf{P}_\lambda$ *is the residual projection matrix and* $\mathbf{X}_{\langle K \rangle}$ *is the regressor matrix with the influence of the differencing matrix* $\mathbf{K}$ *removed.*
Note that the LS estimator of $\boldsymbol{\beta}_\lambda$ *is close to the LS estimator for* $\lambda \to \infty$*. In this case* $\mathbf{P}_\lambda = \mathbf{P}_\infty$ *and has the idempotency property of projectors* $\mathbf{P}_\infty^2 = \mathbf{P}_\infty$ *and* $\mathbf{X}_{\langle K \rangle}$ *is given by* $\mathbf{P}_\infty \mathbf{X}$*. The limiting LS estimate is*

$$\hat{\boldsymbol{\beta}}_\infty = (\mathbf{X}_{\langle K \rangle}^\top \mathbf{X}_{\langle K \rangle})^{-1}\mathbf{X}_{\langle K \rangle}^\top \mathbf{y},$$

which is the usual OLS estimator but with purged regressors $\mathbf{X}_{\langle K \rangle}$*.*

Proof 5. *In case of* $n_2 = 0$*, i.e. in the partial non-informative case, the smoothing result is*

$$\begin{aligned} \boldsymbol{\tau}_{**0} &= \hat{\boldsymbol{\tau}} - \mathbf{A}_{**}^{-1}\mathbf{X}(\mathbf{X}^\top \mathbf{P}_\lambda \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y} \\ &= \hat{\boldsymbol{\tau}} - \mathbf{A}_{**}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_\lambda \\ &= \mathbf{y}_{HP} - \mathbf{y}_{\langle Xb \rangle}, \end{aligned} \tag{45}$$

with $\mathbf{y}_{\langle Xb \rangle} = \mathbf{A}_{**}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_\lambda$ *denotes the component that contains the* $\mathbf{X}$ *regressors and* $\mathbf{G}$ *is*

$$\mathbf{G} = (\mathbf{I}_x - \mathbf{X}^\top \mathbf{A}_{**}^{-1}\mathbf{X})^{-1} = (n_2 \mathbf{I}_p + \mathbf{X}^\top(\mathbf{I}_p - \mathbf{A}_{**}^{-1})\mathbf{X})^{-1} = (n_2 \mathbf{I}_p - \mathbf{X}^\top \mathbf{P}_\lambda \mathbf{X})^{-1} \tag{46}$$

The fully non-informative case can be obtained because $\mathbf{P}_\lambda$ reduces for $\lambda \to \infty$ to

$$\mathbf{P}_\lambda = \mathbf{I}_n - \mathbf{A}_{**}^{-1} \to \mathbf{P}_\infty = (\mathbf{K}^\top(\mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}) \tag{47}$$

and that $\hat{\beta}_\lambda$ can be expressed as

$$\hat{\boldsymbol{\beta}}_\lambda = (\dot{\mathbf{X}}^\top(\mathbf{I}/\lambda + \mathbf{K}\mathbf{K}^\top)^{-1}\dot{\mathbf{X}})^{-1}\dot{\mathbf{X}}^\top(\mathbf{I}/\lambda + \mathbf{K}\mathbf{K}^\top)^{-1}\dot{\mathbf{y}} \tag{48}$$

with $\dot{\mathbf{X}} = \mathbf{K}\mathbf{X}$ and $\dot{\mathbf{y}} = \mathbf{K}\mathbf{y}$. This estimator can be viewed as a generalized ridge estimator (see Hoerl and

Kennard (1970)) since we add $\varepsilon = 1/\lambda$ of the unity matrix to the singular $\mathbf{K}\mathbf{K}^\top$ matrix.

4.2. MCMC for the extended HP (eHP) smoother model

In a Bayesian HP smoothing model we have to proceed in the usual way and specify a prior distribution for the parameters in (29):

$$\mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}], \quad \boldsymbol{\beta} \sim \mathcal{N}[\boldsymbol{\beta}_*, \mathbf{H}_*], \quad \sigma^{-2} \sim \Gamma[\sigma_*^2 n_*/2, n_*/2]. \tag{49}$$

The estimation of the parameters in the extended HP model (29) can be done by a simple MCMC procedure.

Theorem 6. *[MCMC for the extended HP (eHP) model]*

The posterior simulator of the parameters $\theta = (\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^{-2})$ of the extended HP model (29) with prior (49) is given by the following iteration:

1. *Start with $\sigma^2 = \sigma^2_{OLS}$ in the auxiliary model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$;*
2. *Draw $\boldsymbol{\beta}$ from $\mathcal{N}[\boldsymbol{\beta} \mid \boldsymbol{\beta}_{**}, \mathbf{H}_{**}]$;*
3. *Draw $\boldsymbol{\tau}$ from $\mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \mathbf{A}_{**}]$;*
4. *Draw σ^{-2} from $\Gamma[\sigma^{-2} \mid s^2_{**} n_{**}/2, n_{**}/2]$;*
5. *Repeat until convergence.*

The hyper-parameters of the fcd's are given in the proof: (51), (53) and (55).

Proof 6. *The full conditional distributions (fcd) are:*

1. *The fcd for the beta regression coefficients is*

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \mathcal{N}[\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T] \\ &= \mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_{**}, \mathbf{H}_{**}] \end{aligned} \tag{50}$$

with the parameters

$$\begin{aligned} \mathbf{H}_{**}^{-1} &= \mathbf{H}_*^{-1} + \sigma^{-2}\mathbf{X}^\top\mathbf{X}, \\ \mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}_*^{-1}\mathbf{b}_* + \sigma^{-2}\mathbf{X}^\top(\mathbf{y} - \boldsymbol{\tau})] \end{aligned} \tag{51}$$

2. *The fcd for the residual precision σ^{-2}*

$$p(\sigma^{-2} \mid \boldsymbol{\tau}, \mathbf{y}) \propto \Gamma\left[\sigma^{-2} \mid n_{**}s^2_{**}, n_{**}/2\right] \tag{52}$$

we find a gamma distribution with the parameters

$$\begin{aligned} n_{**} &= n_* + n, \\ n_{**}s^2_{**} &= n_* s^2_* + \sum_{i=1}^{n} (y_i - \tau_i - \mathbf{x}_i\boldsymbol{\beta})^2 \end{aligned} \tag{53}$$

3. *The fcd for the* $\boldsymbol{\tau}$ *coefficients is*

$$\begin{aligned} p(\boldsymbol{\tau} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\tau} \mid \mathbf{0}, \mathbf{A}_*] \cdot \mathcal{N}[\mathbf{y} \mid \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T] \\ &= \mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \mathbf{A}_{**}] \end{aligned} \tag{54}$$

with the parameters $\boldsymbol{\tau}_{**} = \mathbf{A}_{**}\mathbf{y}$ *and*

$$\mathbf{A}_{**}^{-1} = \mathbf{A}_*^{-1} + \sigma^{-2}\mathbf{X}^\top \mathbf{X}. \tag{55}$$

5. Application: Smoothing the airline passenger series

For the time series smoothing example we use the airline passenger series from the web site: "Time Series Data Library" of Hyndman (2010). The aim is to show that there is a difference between the smoothed series using the extended HP filter and a simple HP smooth of the residuals. We can remove either a linear or a quadratic trend from the airline passenger series.

First, we look at the regression estimates after fitting the series with a linear or a quadratic trend:

```
lm(formula = y ~ a); Residuals:
      Min        1Q    Median        3Q       Max
-0.134016 -0.045113 -0.007798  0.042291  0.128280
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.4047545  0.0050323  477.87   <2e-16 ***
a           0.0043640  0.0001211   36.05   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.06038 on 142 degrees of freedom
Multiple R-squared: 0.9015,     Adjusted R-squared: 0.9008
F-statistic:  1300 on 1 and 142 DF,  p-value: < 2.2e-16

lm(formula = y ~ a + a2); Residuals:
      Min        1Q    Median        3Q       Max
-0.127143 -0.038184 -0.009388  0.042123  0.122287
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.421e+00  7.343e-03 329.717  < 2e-16 ***
a            4.373e-03  1.178e-04  37.122  < 2e-16 ***
a2          -9.515e-06  3.168e-06  -3.004  0.00316 **
---
Signif. codes:  0 *** 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.05875 on 141 degrees of freedom
Multiple R-squared: 0.9074,     Adjusted R-squared: 0.9061
F-statistic:   691 on 2 and 141 DF,  p-value: < 2.2e-16
```

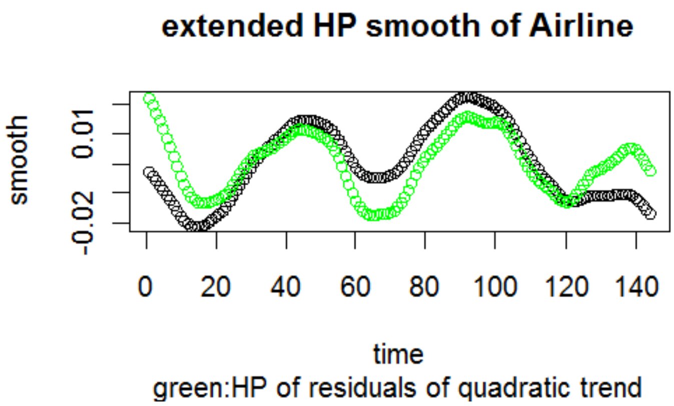
The eHP filter of model (34) is shown in Figure 2 where we compare the smoothed series with the simple HP smooth after having removed the quadratic trend from the airline passenger series. We see that the peaks of the smooth are at the same positions but the extended HP smooth produces a smoother "HP smooth".

6. A Spatial HP smoothness procedure

In analogy to the HP filter for time series models we consider a spatial HP filter model based on a spatial autoregression (SAR) model of first order, which is defined as (see Anselin 1988)

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad with \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_n], \tag{56}$$

Figure 1: eHP smooth of airline passengers with quadratic trend



where $\mathbf{W}$ is a row-normalized weight matrix, $\mathbf{Wy}$ is the first order spatial lag of $\mathbf{y}$, and ρ is the spatial correlation coefficient (see Lesage and Pace 2009). Model (56) can be viewed as a SAR(1) model is equivalent to the transformed model

$$\mathbf{Ry} = \boldsymbol{\tau} + \varepsilon, \quad or \quad \mathbf{y} \sim \mathcal{N}[\mathbf{R}^{-1}\boldsymbol{\tau}, \sigma^2(\mathbf{R}^\top\mathbf{R})^{-1}]$$

with the spatial spread matrix $\mathbf{R} = \mathbf{I}_n - \rho\mathbf{W}$.

Using the SSL principle (1) we can define a spatial HP-type smoothness filter. We assume a HP smoothing model based on a SAR(1) model

$$\mathbf{y} \sim \mathcal{N}[\rho\mathbf{Wy} + \boldsymbol{\tau}, \sigma^2\mathbf{I}_n] \quad or \quad \mathbf{y} \sim \mathcal{N}[\mathbf{R}^{-1}\boldsymbol{\tau}, \sigma^2(\mathbf{R}^\top\mathbf{R})^{-1}] \tag{57}$$

with the spread matrix $\mathbf{R} = (\mathbf{I}_n - \rho\mathbf{Wy})$.

For the HP-type smoothing problem in space we have to define a metric: what is a first and second order spatial difference? For the nearest neighbors (NN) metric this is easy: the first order is the difference to the first NN and the second order is the difference to the second order NN. In analogy to the HP filter (3) for time series we can write the spatial HP-type smoothing problem as the minimizer of the SSL function as in Definition 1

$$\begin{aligned} \boldsymbol{\tau}_{**} &= \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad with \\ SSL(\boldsymbol{\tau}) &= (\mathbf{Ry} - \boldsymbol{\tau})^\top(\mathbf{Ry} - \boldsymbol{\tau}) + \lambda\sum_{i=1}^{n}(\mathrm{w}\,_i^{(0)}\mathbf{y} - 2\mathrm{w}\,_i^{(1)}\mathbf{y} - \mathrm{w}\,_i^{(2)}\mathbf{y})^2. \end{aligned} \tag{58}$$

The idea is that the penalty term minimizes the second order smoothness, i.e. the local distance between the first 3 neighbors and the current observation, which in the spatial context is reflected by the original observation $\mathrm{W}^{(0)} = \mathbf{I}_n$, the first order $\mathrm{W}^{(1)}$ and second order $\mathrm{W}^{(2)}$ NN weighting matrix:

$$\begin{aligned} smooth &= \sum_{i=1}^{n}(y_i - \mathrm{w}\,_i^{(1)}\mathbf{y} - \mathrm{w}\,_i^{(1)}\mathbf{y} + \mathrm{w}\,_i^{(2)}\mathbf{y})^2 \\ &= \sum_{i=1}^{n}\Delta^{(2)}\mathrm{w}\,_i\mathbf{y} = \mathbf{y}^\top\mathbf{K}^\top\mathbf{K}\mathbf{y} \end{aligned} \tag{59}$$

with $\mathrm{w}\,_i^{(1)}$, and $\mathrm{w}\,_i^{(2)}$ being the i-th row of the first, and second order NN weighting matrices $\mathrm{W}^{(1)}$ and

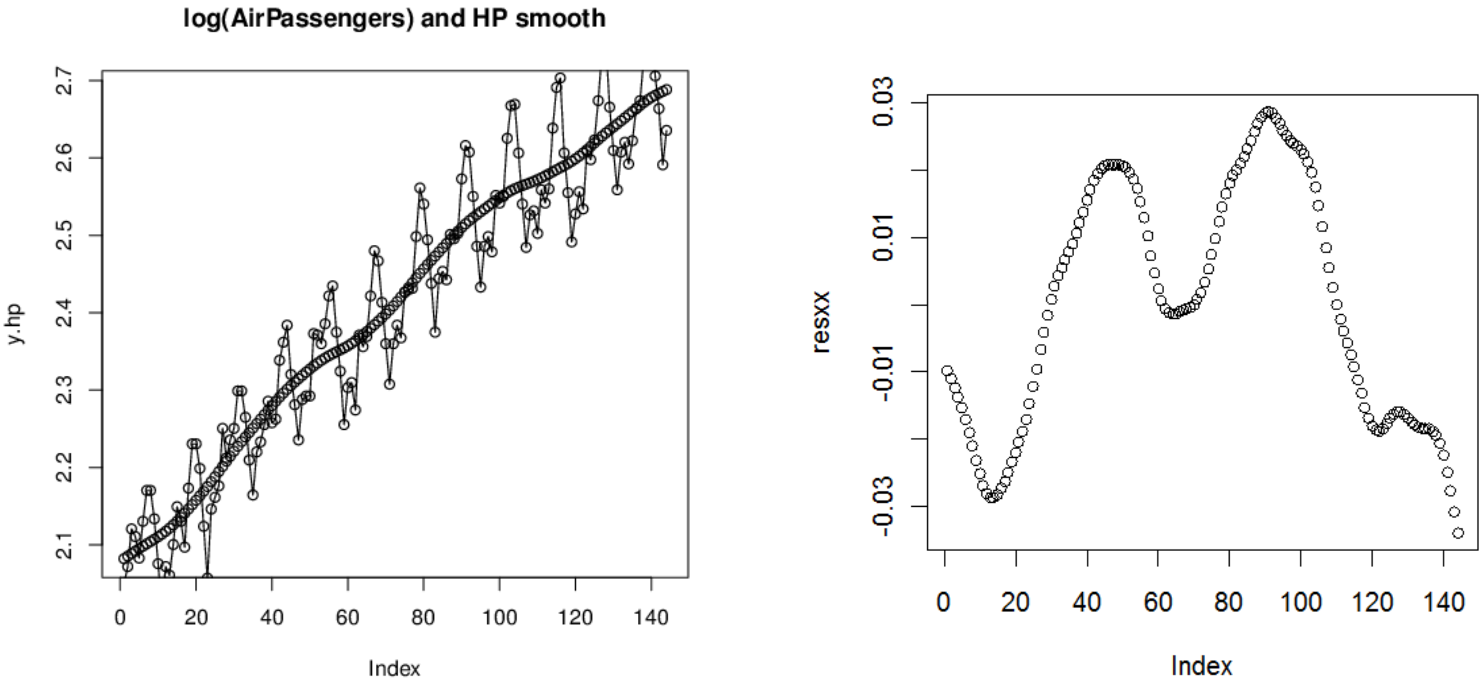


Figure 2: HP smooth of log airline passengers HP smooth of linear airline passenger trend

$\mathrm{W}^{(2)}$, respectively, and the second order differencing matrix is $\Delta^{(2)}\mathrm{w}_{i}\mathbf{y} = \Delta\mathrm{w}_{i}\mathbf{y}^{(1)} - \Delta\mathrm{w}_{i}\mathbf{y}^{(2)}$ with the neighborhood matrix W being partitioned row-wise: $\mathrm{W} = \begin{pmatrix} \mathrm{w}_1 \\ \dots \\ \mathrm{w}_n \end{pmatrix}$.

This means that the spatial HP filter $\boldsymbol{\tau}$ minimizes the SSL function in (1) using a spatial smooth penalty function. The error sum of squares is $ESS(\boldsymbol{\tau}) = \sum_{i=1}^{n}(y_i - \tau_i)^2$ between the HP smoother τ_i and the observations y_i's while the spatial penalty term is defined in (59).

The spatial differencing matrix $\mathbf{K}$ is of order $n \times n$, since we do not lose observations in the differencing process, which has the following form:

$$\mathbf{K} = \begin{pmatrix} \mathrm{w}_1^{(0)} & -2\mathrm{w}_1^{(1)} & \mathrm{w}_1^{(2)} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathrm{w}_2^{(0)} & -2\mathrm{w}_2^{(1)} & \mathrm{w}_2^{(2)} & 0 & \dots & 0 & 0 & 0 \\ & \dots & & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & 0 & 0 & \dots & w_n^{(0)} & -2\mathrm{w}_n^{(1)} & \mathrm{w}_n^{(2)} \end{pmatrix} (\mathbf{1}_n \otimes \mathbf{I}_n) = \begin{pmatrix} \mathrm{w}_1^{(0)} - 2\mathrm{w}_1^{(1)} + \mathrm{w}_1^{(2)} \\ \mathrm{w}_2^{(0)} - 2\mathrm{w}_2^{(1)} + \mathrm{w}_2^{(2)} \\ \dots \\ \mathrm{w}_n^{(0)} - 2\mathrm{w}_n^{(1)} + \mathrm{w}_n^{(2)} \end{pmatrix} : n \times n \tag{60}$$

The $n^2 \times n$ block matrix $(\mathbf{1}_n \otimes \mathbf{I}_n)$ is a block row summation operator for the spatial differencing matrix, adding up the $\mathrm{w}_i^{(d)}$ terms. Now we can formulate a HP smoother for spatial cross-section in similar way as in Theorem 1.

Theorem 7. *[The spatial HP Filter] We consider the SAR model (57) and the spatial smoothness prior*

(59) based on distances.

The minimum of the spatial HP-type smoothing problem using the SSL (smoothed squared loss) principle in (1) is obtained by minimizing the quadratic form in $\boldsymbol{\tau}$, where we rewrite (3) with $\mathbf{y} = [y_1, ..., y_n]'$, $\boldsymbol{\tau} = [\tau_1, ..., \tau_n]^\top$ and the second order differencing matrix $\mathbf{K} : n \times n$, defined in (60), as

$$\min_{\boldsymbol{\tau}} (\mathbf{Ry} - \boldsymbol{\tau})^\top (\mathbf{Ry} - \boldsymbol{\tau}) + \lambda \boldsymbol{\tau}^\top \mathbf{K}^\top \mathbf{K} \boldsymbol{\tau} \tag{61}$$

is attained at the posterior mean (the "least squares estimate under restrictions") and is the solution to the optimisation problem:

$$\boldsymbol{\tau}_{**} = [\mathbf{R}^\top \mathbf{R} + \lambda \mathbf{K}^\top \mathbf{K}]^{-1} \mathbf{Ry}. \tag{62}$$

with $\mathbf{R} = \mathbf{I}_n - \rho \mathbf{W}$. Since $\boldsymbol{\tau}_{**}$ (sometimes denoted also by $\hat{\tau}$ to emphasize the posterior mean as an estimate) depends on the unknown ρ, we have to minimize the variance matrix of $\boldsymbol{\tau}_{**}$ with respect to ρ. The variance of the posterior mean is $Var(\boldsymbol{\tau}_{**}) = [\mathbf{R}^\top \mathbf{R} + \lambda \mathbf{K}^\top \mathbf{K}]^{-1}$.

Proof 7. *The proof relies on rewriting the optimisation problem as a sum of 2 quadratic forms in τ and to apply Theorem 9 of the appendix:*

$$(\mathbf{Ry} - \boldsymbol{\tau})^\top (\mathbf{Ry} - \boldsymbol{\tau}) + \lambda \boldsymbol{\tau}^\top \mathbf{K}^\top \mathbf{K} \boldsymbol{\tau} = (\boldsymbol{\tau} - \boldsymbol{\tau}_{**})^\top (\boldsymbol{\tau} - \boldsymbol{\tau}_{**}) + \mathbf{y}^\top \lambda \mathbf{K}^\top \mathbf{K} (\lambda \mathbf{K}^\top \mathbf{K} + \mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \mathbf{Ry} \tag{63}$$

with the posterior mean $\boldsymbol{\tau}_{**} = \mathbf{A}_{**}^{-1} \mathbf{R}^\top \mathbf{y}$ and the posterior precision matrix

$$\mathbf{A}_{**} = [\mathbf{R}^\top \mathbf{R} + \lambda \mathbf{K}^\top \mathbf{K}].$$

Finally, the point predictor for the spatial HP smooth is given by the posterior mean $\boldsymbol{\tau}_{**}$.

Theorem 8. *[The HP filter as posterior mean] The minimizer of the SSL minimisation problem in definition 1 under the assumption of a normal distribution for the idem-parameterized regression model[7] and the stochastic smoothness model is the posterior mean $\hat{\mathbf{y}} = \boldsymbol{\tau}_{**}$, given by*

$$\boldsymbol{\tau}_{**} = (\mathbf{I}_n + \lambda \mathbf{K}^\top \mathbf{K})^{-1} \mathbf{y}. \tag{64}$$

Proof 8. *We combine both, the quadratic form in $\boldsymbol{\tau}$ of the idem-parameterized regression model and the stochastic smoothness model using theorem (9). The result is a function of a single quadratic form $\boldsymbol{\tau}$ that is minimized by the mean of the quadratic form that corresponds to the posterior mean of the equivalent Bayesian regression model.*

7. Applications of spatial HP filtering

In this section we show how the spatial HP model can be applied to smooth the regional GDP across the 239 (contiguous) NUTS-2 regions in Europe for the year 2005. The data with the coordinates of the center points of the NUTS-2 regions are taken from EUROSTAT.

To define a smooth surface for a spatial cross-sectional data set we have to define a differencing matrix. As it was shown in the above section, this can be easily done if we have a distance matrix between the centers of the NUTS-2 regions. Thus we identify for each region a nearest neighbor (by distance) and a second nearest neighbor (also by distance). This produces the following **K** matrix, where - for demonstration - we display the first 6 rows.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]
[1,]	1	0	-2	0	0	0	0	0	0	0	0	0	0	0	0
[2,]	1	1	-2	0	0	0	0	0	0	0	0	0	0	0	0
[3,]	-2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
[4,]	0	0	0	1	-2	0	1	0	0	0	0	0	0	0	0
[5,]	1	0	0	-2	1	0	0	0	0	0	0	0	0	0	0
[6,]	0	-2	0	0	0	1	1	0	0	0	0	0	0	0	0

The effect of the spatial smoothing is seen in alphabetical order of the 27 countries[8] in Figure 3. The volatility of the smooth can be attributed to the heterogeneity of the countries and the volatility within countries.

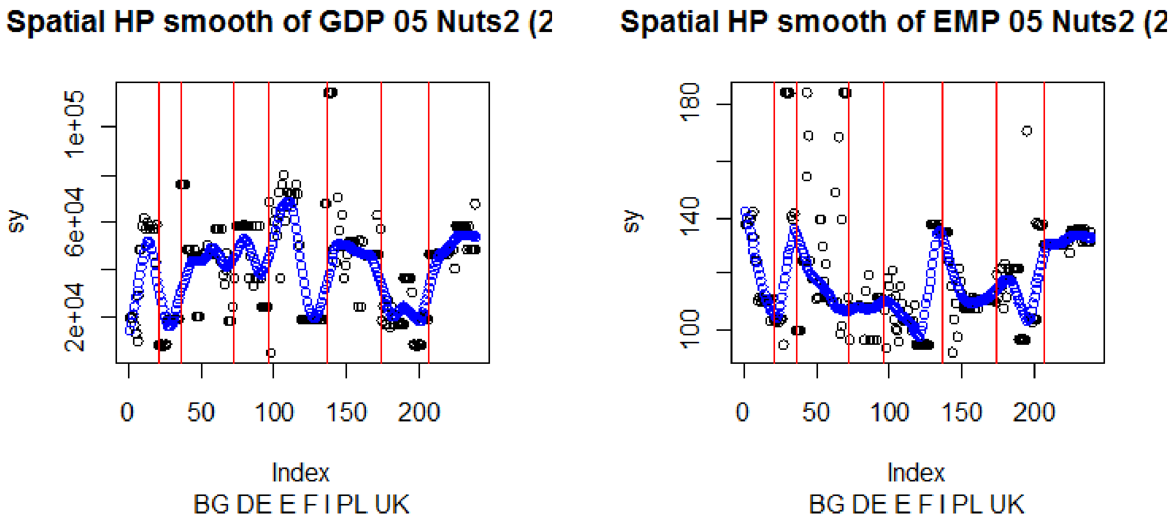
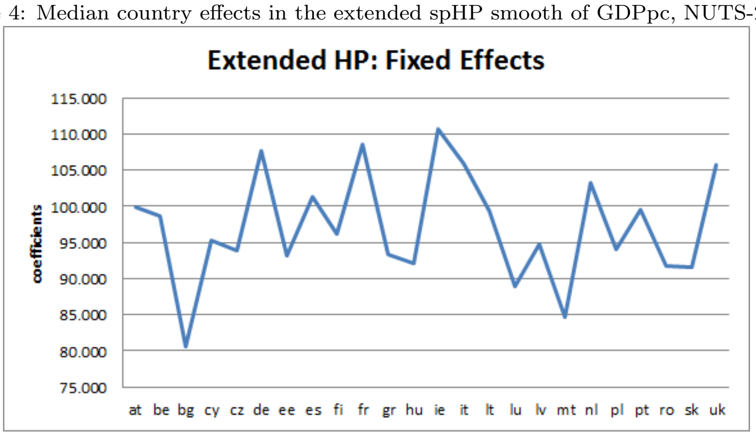


Figure 3: Spatial HP smooth of GDP 05, NUTS-2, 2005 Spatial HP smooth of Employment, NUTS-2, 2005

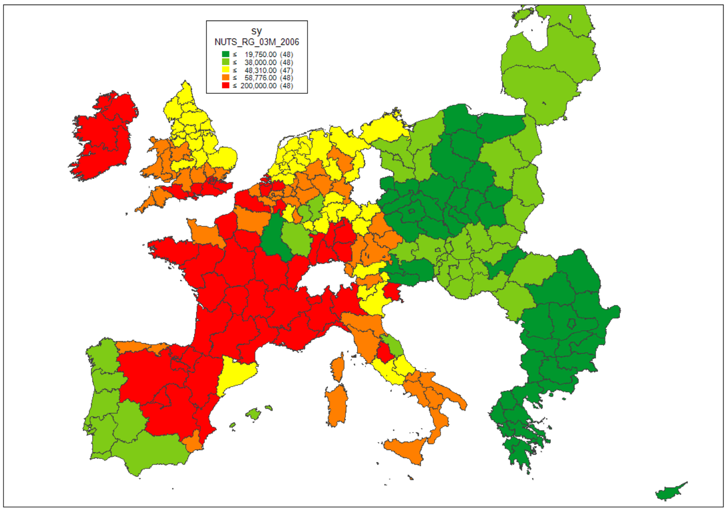
The median effects of the **X** matrix in the extended spatial HP procedure estimated with MCMC are shown in Figure 4. In our case these are the median effects of the 25 country dummy variables (or fixed effects): The smallest one is Portugal and the largest one is Malta.

Figure 4: Median country effects in the extended spHP smooth of GDPpc, NUTS-2, 2005



[8] AT BE BG CY CZ DE EE E FI F GR HU IE I LT LU LV MT NL PL PT RO SK UK

The geographical maps for the smoothed GDP and GDPpc of NUTS-2 regions are given in the Figure 5 and in Figure 6, respectively, together with the observed raw values.



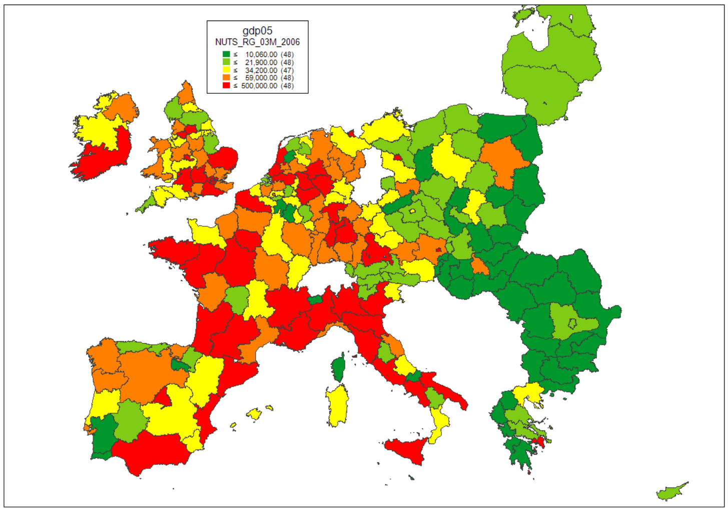
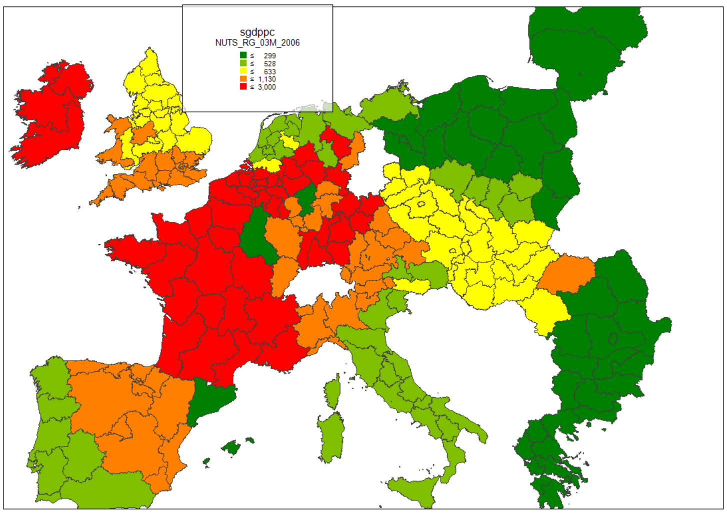


Figure 5: Spatial HP smooth of GDP NUTS-2, 2005 Map of 239 GDP NUTS-2 regions, 2005 (raw data)



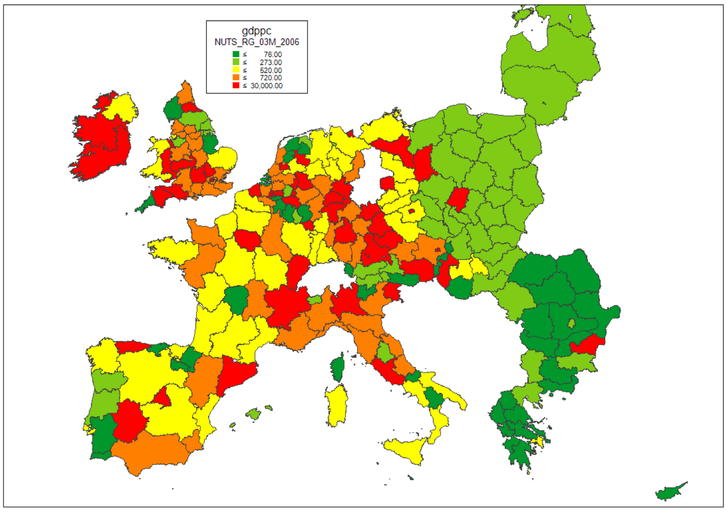


Figure 6: Spatial HP smooth of GDPpc, NUTS-2, 2005 239 GDPpc NUTS-2 regions, 2005 (raw data)

7.1. Spatial smoothing: the results for 239 European regions

We consider the GDP and employment data for 239 NUTS-2 regions in Europe for the year 2005. Some islands and oversee regions were left out, because the distance measure for the spatial lags used are car driving times between the centers of the regions. (These were obtained by own calculations based on pairwise queries by internet search machines.)

```
lm(formula = log(y) ~ 0 + ZZ)
Residuals:
     Min       1Q   Median       3Q      Max
```

```
-3.00630 -0.40641 -0.02213  0.46751  2.22527
Coefficients:
       Estimate Std. Error t value Pr(>|t|)
Dagg   1.4260     0.6161   2.32   0.0216 *
at     9.9837     0.2815  35.47  < 0.000 ***
be     9.8617     0.2607  37.83  < 0.000 ***
bg     8.0539     0.3448  23.36  < 0.000 ***
cy     9.5222     0.8445  11.28  < 0.000 ***
cz     9.3844     0.2986  31.43  < 0.000 ***
de    10.7655     0.1407  76.49  < 0.000 ***
ee     9.3245     0.8445  11.04  < 0.000 ***
es    10.1280     0.1937  52.28  < 0.000 ***
fi     9.6111     0.3777  25.45  < 0.000 ***
fr    10.8617     0.1800  60.33  < 0.000 ***
gr     9.3272     0.2815  33.14  < 0.000 ***
hu     9.2109     0.3192  28.86  < 0.000 ***
ie    11.0647     0.5971  18.53  < 0.000 ***
it    10.5937     0.1843  57.49  < 0.000 ***
lt     9.9366     0.8445  11.77  < 0.000 ***
lu     8.8840     1.0453   8.50    0.000 ***
lv     9.4736     0.8445  11.22  < 0.000 ***
mt     8.4671     0.8445  10.03  < 0.000 ***
nl    10.3268     0.2438  42.36  < 0.000 ***
pl     9.4063     0.2111  44.55  < 0.000 ***
pt     9.9515     0.3777  26.35  < 0.000 ***
ro     9.1704     0.2986  30.72  < 0.000 ***
sk     9.1493     0.4222  21.67  < 0.000 ***
uk    10.5717     0.1482  71.34  < 0.000 ***
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 '.' 0.1 '' 1

Residual standard error: 0.8445 on 214 degrees of freedom
Multiple R-squared: 0.9939,     Adjusted R-squared: 0.9931
F-statistic:  1386 on 25 and 214 df,  p-value: < 2.2e-16

Ordered effects:
 Dagg      bg     mt     lu     sk     ro     hu    ee
 1.426  8.054  8.467  8.884  9.149  9.170  9.211  9.325
  gr      cz     pl     lv     cy     fi     be    lt
 9.327  9.384  9.406  9.474  9.522  9.611  9.862  9.937
  pt      at     es     nl     uk     it     de     fr    ie
 9.952  9.984 10.128 10.327 10.571 10.594 10.766 10.862  11.065
```

7.2. The spatial extended HP (eHP) smoother

In this section we apply the spatial HP smoother to the following extended smoothing problem: we correct the spatial cross section of GDP and log GDP first with the fixed effect dummies for the 25 countries involved and then apply the spatial HP smoother to the "purged" cross-sectional observations. The results of the extended spatial smoother (with $\lambda = 1600$) can be seen in the next figures.

7.3. Employment

In this section we report the spatial smoothing results for the regional employment data in 2005.

Figure 8 shows the raw data together with the smooth of the emplozment data in 2005: the first things to note are the high employment effects in central Poland and Romania. The smooth in Figure 8 shows the smooth (posterior mean) of the spatial HP model while Figure 9 shows the smooth (posterior mean) of the spatial extended HP model. The $\mathbf{X}$ matrix of the extended model (eHP) just contains the fixed effect dummy variables for the countries plus an extra dummy for the new central and eastern European states (CEE). The border of the regions in the East and West of the smooth can be seen in both figures, which stretch until France. The somewhat unexpected map is due to the fact that German regions have less employment than the regions in Poland and Romania. Therefore we see higher smoothed values at the periphery and lower values in the center (Germany, the Czech Republic and Austria.) Also, by taking into account the large variation of levels across EU countries we see that these "low smooth" values are still present in those 3 central European states.

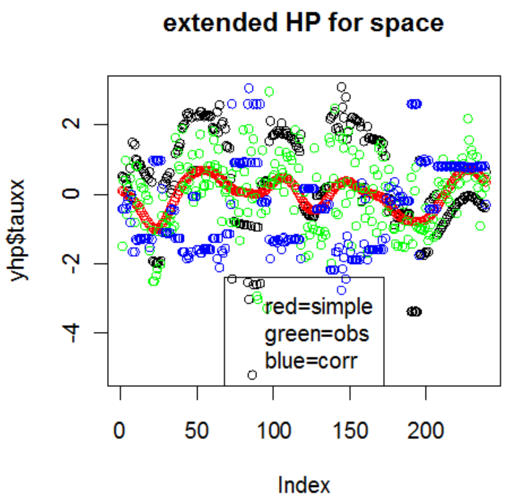
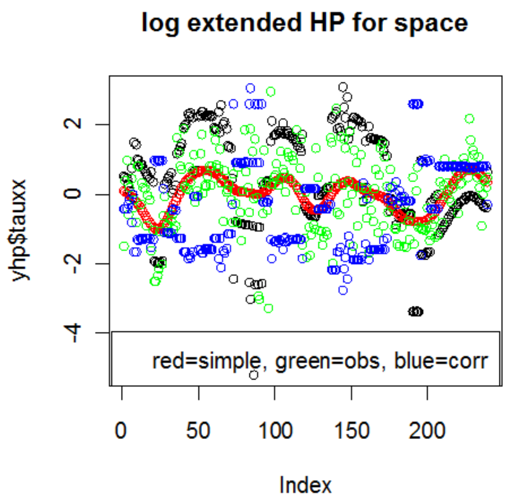


Figure 7: Spatial extended HP smooth of GDP Nuts-2, 2005



Spatial extended HP smooth of log GDP Nuts-2, 2005

8. Summary

This paper has shown how the HP filter can be viewed from a Bayesian point of view and how this procedure can be extended to an extended HP filter model with additional regressor variables in a time series context and a newly proposed spatial HP filter for cross-sectional data. In the time series context the new approach leads to the extended HP smoother that allows to incorporate other factors and regressors into the smoothing problem, which leads to a purge of the target data to be smoothed from these other factors, like trend, outliers and fixed effects. This extension of the HP filter model was demonstrated for the time series case using the well-known airline passenger data and shows how the trend can be removed succesfully from the data before the smoothing procedure produces the final result.

The Bayesian view of the HP smoothing problem allows an easy interpretation of the smoothing constant: it is the hypothetical sample size of the prior information that is used in the HP smoothing model. To produce a smooth output one has to increase the prior precision to stick quite close to the chosen "smooth" prior, which is defined by the second difference of the smooth component, i.e. the parameter vector to be estimated. In the extended HP model we have to split up this hypothetical sample size of the prior into the two parts of the model: The smooth part needs a high precision parameter to stick close to the prior and to produce the HP-type of smooth, while the regression part defines the extended part of the smoothing model and needs the (usual low) precision parameter if we want a flexible fit to the other regressor variables.

In the spatial context, the extended HP filter allows a spatial smoothing of data and this was demonstrated for the 239 NUTS-2 regions of the European Union for GDP and employment data. The smoothness in a spatial context is defined by the distance of neighboring regions. The spatial extended HP smoother can be computed easily using MCMC procedures of the linear regression model or the spatial autoregression (SAR) model. It is argued that this new family of extended HP procedures opens a new approach for smoothing output variables in more complex models that requires more adjustments and simplifications before the smoothing can be done, and the Bayesian interpretation shows to give more flexibility for the prior information that combines the smooth and the non-smooth part in such more complex HP-type smoothing models. Thus, our approach has demonstrated that econometric smoothing problems can be either embedded in simple univariate set-ups or in complex model-based applications.

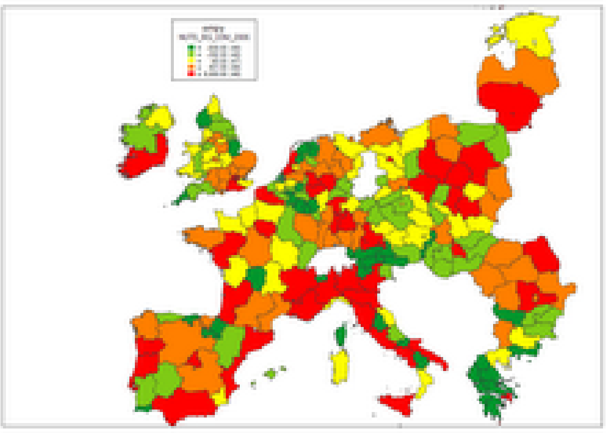
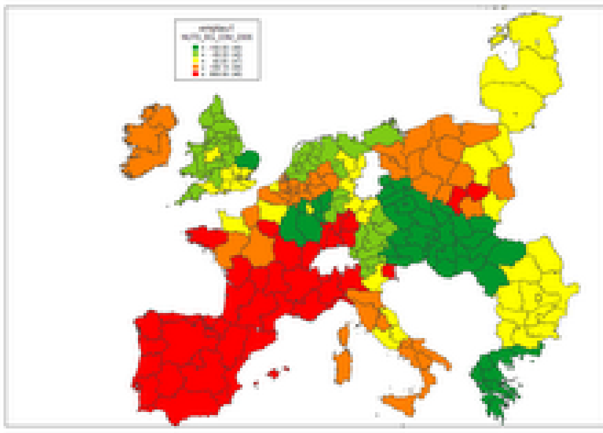
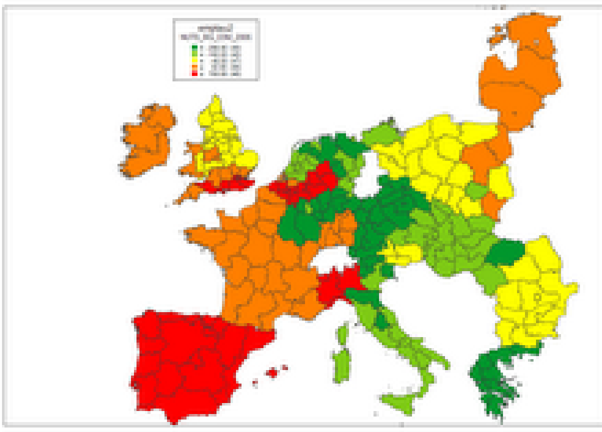


Figure 8: Employment: NUTS-2, 2005 (raw data)



MCMC Spatial HP smooth of Employment NUTS-2, 2005

Figure 9: MCMC of the spatial extended HP model, smooth of Employment NUTS-2, 2005



9. References

Anselin, L. (1988), Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, The Netherlands.

Anselin, L. and R. Florax (1995), New Directions in Spatial Econometrics, Springer-Verlag, Berlin.

Baxter, Marianne, and Robert G. King (1999), Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series, Review of Economics and Statistics, vol.81(4), 575-93

Blackburn, Keith, and Morten O. Ravn (1991), Contemporary Macroeconomic Fluctuations: an International Perspective, Discussion Paper 1991-06, University of Southampton.

Blackburn, Keith, and Morten O. Ravn (1992), Business Cycles in the UK: Facts and Fictions, Economica, vol.59, 383-401.

Cressie, N. (1993) Statistics for Spatial Data, Wiley, New York.

Danthine, J.P. and Girardin M. (1989), Business Cycles in Switzerland: A Comparative Study, European Economic Review 33, 31-50.

Harvey, Andrew C. (1993), Forecasting Structural Time Series and the Kalman Filter, Cambridge UK: Cambridge University Press.

Hodrick, Robert J. and Edward C. Prescott (1980), Postwar U.S. Business Cycles: an Empirical Investigation, Discussion Paper no. 451, Carnegie Mellon University. This paper was printed in an updated version as: Postwar U.S. Business Cycles: an Empirical Investigation, Journal of Money, Credit and Banking, 1997, vol.29 (1), 1-16.

Hodrick, R. and E.P. Prescott (1997), Postwar Business Cycles: An Empirical Investigation, Journal of Money, Credit, and Banking, 29, 1-16.

Hoerl, A. E. and Kennard, R. W. (1970), Ridge Regression: Biased Estimation for Non-orthogonal Problems. Technometrics, 12, 55-67.

Hyndman, R.J. (2010), Time Series Data Library, //http://robjhyndman.com/TSDL. Accessed on November 2010.

Kaiser, Regina and Agostin Maravall (1999), Estimation of the Business Cycle: a modified Hodrick-Prescott filter, Spanish Economic Review, 1 (2), 175-206.

Kim, K.H., R.A. Buckle, and V.B. Hall (1994), Key Features of New Zealand Business Cycles, Economic Record, vol.70(208), 56-72.

King, Robert G. and Sergio T. Rebelo (1993), Low Frequency Filtering and Real Business Cycles, Journal of Economic Dynamics and Control vol. 17, no. 1-2, 207-231.

Kydland, Finn E. and Edward C. Prescott (1990), Business Cycles: Real Facts and a Monetary Myth, Federal Reserve Bank of Minneapolis Quarterly Review vol.14, 3-18.

Luetkepohl H. (1991), Introduction to Multiple Time Series Analysis, Springer, New York.

Ord, J.K. (1975), Estimation Methods for Models of Spatial Interaction, J. Amer. Statist. Assoc., 70, 120-126.

Paelinck, J., and L. Klaassen (1979), Spatial Econometrics, Saxon House, Farnborough.

Polasek W. and R. Sellner (2011), Does Globalization affect Regional Growth? Evidence for EU27 NUTS-2 Regions, IHS-Vienna, discussion paper.

Ravn, Morten O. (1997), International Business Cycles in Theory and in Practise, Journal of International Money and Finance, vol.16(2), 255-83.

Ravn, Morten O., and Harald Uhlig (2002), On Adjusting the HP-Filter for the Frequency of Observations, Review of Economics and Statistics, vol. 84(2), 371-76.

Acknowledgement: I want to thank R. Sellner for helping to draw the geographic maps and the data set that was prepared for a previous joint paper, see Polasek and Sellner (2011).

APPENDIX A

10. Results on Combination of Quadratic Forms

We list some standard results of normal Bayes models for combining normal densities:

Theorem 9 (Combination of Quadratic Forms).

Let $\mathbf{H}$ *and* $\mathbf{H}_*$ *be two symmetric quadratic matrices. Then the sum of the two quadratic forms can be combined as*

$$(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{H}(\boldsymbol{\beta}-\mathbf{b})+(\boldsymbol{\beta}-\mathbf{b}_*)^\top\mathbf{H}_*(\boldsymbol{\beta}-\mathbf{b}_*)$$
$$=(\boldsymbol{\beta}-\mathbf{b}_{**})^\top\mathbf{H}_{**}(\boldsymbol{\beta}-\mathbf{b}_{**})+(\mathbf{b}-\mathbf{b}_*)^\top\mathbf{H}_*(\mathbf{H}_*+\mathbf{H})^{-1}\mathbf{H}(\mathbf{b}-\mathbf{b}_*)$$

with the parameters

$$\begin{aligned}\mathbf{H}_{**} &= \mathbf{H}_*+\mathbf{H},\\ \mathbf{b}_{**} &= \mathbf{H}_{**}^{-1}(\mathbf{H}_*\mathbf{b}_*+\mathbf{Hb}).\end{aligned} \tag{65}$$

Lemma 1 (MESS-Decomposition for regression models). *Let* $\boldsymbol{\beta}$ *be any regression vector in the model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ *and* $\mathbf{b}$ *the OLS estimator* $\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$. *Then the residual sum of squares can be decomposed as*

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = n\hat{\sigma}^2 + (\boldsymbol{\beta} - \mathbf{b})^\top\mathbf{X}^\top\mathbf{X}(\boldsymbol{\beta} - \mathbf{b}) \tag{66}$$

where the error sum of square (ESS) is

$$n\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^\top(\mathbf{y} - \mathbf{X}\mathbf{b}),$$

which is the minimum error sum of squares (and is briefly named MESS) or minimum MSE (MMSE).