The Hodrick-Prescott (HP) filter as a Bayesian regression model

Wolfgang Polasek[a]

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

Abstract

The Hodrick-Prescott (HP) method is a popular smoothing method for economic time series to get a smooth or long-term component of stationary series like growth rates. We show that the HP smoother can be viewed as a Bayesian linear model with a strong prior using differencing matrices for the smoothness component. The HP smoothing approach requires a linear regression model with a Bayesian conjugate multi-normal-gamma distribution. The Bayesian approach also allows to make predictions of the HP smoother on both ends of the time series. Furthermore, we show how Bayes tests can determine the order of smoothness in the HP smoothing model. The extended HP smoothing approach is demonstrated for the non-stationary (textbook) airline passenger time series. Thus, the Bayesian extension of the HP model defines a new class of model-based smoothers for (non-stationary) time series and spatial models.

Keywords: Hodrick-Prescott (HP) smoothers, model selection by marginal likelihoods, multi-normal-gamma distribution, Spatial sales growth data, Bayesian econometrics .

JEL classification: C11, C15, C52, E17, R12 .

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

1. Introduction

Data smoothing in time and space is an important tool for model building. Therefore the understanding of methods should be beyond mechanical applications of black box methods. We will demonstrate in this paper that the extension of the Hodrick-Prescott (HP) smoother can serve as such a role model for smoothing data in time and space. The first approach of this type of 'HP'-smoothing was derived in Leser (1961).

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. For this purpose we have to define the 'multi-normal-gamma' (mNG) family of conjugate distributions. Using the smoothed squared loss (SSL) function, the classical approach to HP smoothing is reviewed in section 2 and the Bayesian embedding into a regression model is explained in section 3. Section 4 describes model selection from a Bayesian perspective using marginal likelihoods and Bayes factors. A final section concludes. The appendix contains a result on combination of quadratic forms and The R program.

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 (briefly 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 as a basis to analyse the turning points in business cycles and many researchers compare their results with those obtained for the US data.

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 minimization 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 $\lambda = 1600$."

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 homolog (i.e. having an equal number of observations and location parameters, yielding actually to an over-parameterized or 'pera'-parametric (from the Greek pera= over) model) regression problem for the observations $\mathbf{y} = [y_1, ..., y_T]'$. This model for obtaining the smooth of a time series under quadratic loss is called in this paper the 'HP regression model'.

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

In this regression model with identity regressor matrix $\mathbf{X} = \mathbf{I}_T$, the HP smoother is defined as parameter vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_T]^\top$ and the 'HP smooth' is the estimated $\boldsymbol{\tau}$ vector. The classical estimation approach for this problem is based on an optimization of a special loss function, which we will call the "smoothed squared loss (SSL) 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* $\hat{\mathbf{y}}$*:*

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

where ESS is an error sum of squares function of the homolog (= 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 assumed to be 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: $SSL = ESS \ + \ \lambda * \texttt{smooth}$ 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 SSL minimization problem is given by the next theorem.

Theorem 1 (The HP smoother as a posterior mean).

We consider the HP smoothing problem in the regression model (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 (or sometimes called the "least squares estimate under restrictions") 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 covariance matrix

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

The second order[1] *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* $SSL = ESS + \lambda * \texttt{smooth}$ *as a sum of 2 quadratic forms in* $\boldsymbol{\tau}$*:*

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

and we apply Theorem 7 of the appendix:

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

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

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

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* $\boldsymbol{\tau}_{**}$*. Therefore the HP smoother the equivalent to a Bayesian normal (homoskedastic) regression model with higly informative prior:*

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

The next theorem summarizes some basic properties of the HP smoother and its non-orthogonal decomposition by 'pre-jectors'[2].

Theorem 2 (Properties of the HP smoother).

For the HP smoother (5) we find the following properties:

1. *The HP 'smooth prejector' is not an orthogonal but skewed projector since it can be decomposed in a similar way as orthogonal projectors* $\mathbf{A}_{**} = (\mathbf{I}_T + \mathbf{K}^\top\lambda\mathbf{K})^{-1} = \mathbf{I}_T - \mathbf{P}_\lambda$ *with the 'rough'-prejector of the smooth*
$$\mathbf{P}_\lambda = \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}. \tag{12}$$
2. *The HP smoother produces the (non-orthogonal) data decomposition*
$$\mathbf{y} = \mathbf{y}_{**} + \hat{\mathbf{e}}. \tag{13}$$
3. *The HP smoother is 'unbiased' in the mean, i.e.* $\bar{\mathbf{y}} = \bar{\mathbf{y}}_{**}$ *since* $Ave(\hat{\mathbf{e}}) = 0$.

[1] Note that the second or higher order differencing matrices can be created from the first order differencing matrix by matrix powers: the second order by $\mathbf{K}_2 = \mathbf{K}_1\mathbf{K}_1$, the p-th order by $\mathbf{K}_p = \mathbf{K}_1^p$.

[2] Indicating a pre-stage of a projection mapping.

Proof 2. *The posterior covariance matrix of the HP smoother* $\mathbf{A}_{**}$ *can be viewed as a smoothing operator since it produces the smooth* $\mathbf{y}_{**} = \mathbf{A}_{**}\mathbf{y}$. *The inverse* $\mathbf{A}_{**}^{-1} = \mathbf{I}_T + \mathbf{K}^\top\lambda\mathbf{K} = \mathbf{A}_\lambda$ *is a linear function of* λ *and can be decomposed by the inversion lemma*[3]

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

from where we obtain (12) and the smoothing pre-jectors of this decomposition satisfy the matrix identity $\mathbf{A}_{**} + \mathbf{P}_\lambda = \mathbf{I}_T$.

Therefore a faster computation of the 'rough-prejector' is $\mathbf{P}_\lambda = \mathbf{I}_T - \mathbf{A}_{**}$. *For the HP data smoother in (5) we find*

$$\begin{aligned} \mathbf{y}_{**} &= (\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K})^{-1}\mathbf{y} = [\mathbf{I}_T - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}]\mathbf{y} \\ &= \mathbf{y} - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K}\mathbf{y} = \mathbf{y} - \hat{\mathbf{e}}. \end{aligned} \tag{15}$$

The second term produces the estimated residual $\hat{\mathbf{e}} = \mathbf{P}_\lambda\mathbf{y}$ *and estimates the rough or noise component* $\hat{\mathbf{e}}$ *of this HP smoothness problem that leads to the well-known data decomposition (13)*

$$data \ = fit \ + \ rough.$$

A simple measure for the amount of smoothing is the variance of the rough: $Var(\hat{\mathbf{e}}) = \sum_t \hat{e}_t^2/T$ or the noise-to-signal ratio $Var(\hat{\mathbf{e}})/Var(\mathbf{y})$ since we find the relative variance decomposition

$$Var(\mathbf{y}_{**})/Var(\mathbf{y}) + Var(\hat{\mathbf{e}})/Var(\mathbf{y}) = 1.$$

Note that the mean of $\hat{\mathbf{e}}$ is zero since $\mathbf{K}\mathbf{1}_T = \mathbf{1}_T^\top\mathbf{K}^\top = \mathbf{0}$ and therefore we have the property $\bar{\mathbf{y}} = \bar{\mathbf{y}}_{**}$, which is also valid for least squares (LS) decompositions.

3. The HP filter as a Bayesian smoothness regression model

The Bayesian HP type smoothing model starts also from the HP type regression (= decomposition) model (1)

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{I}_T], \tag{16}$$

with the identity matrix as "regressors" and where $\boldsymbol{\tau} : T \times 1$ is the equal-sized (homolog) 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 $\mathbf{K} : (T-2) \times T$:

$$\mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}]. \tag{17}$$

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

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

$$\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)(\mathbf{K}^\top\mathbf{K})^{-1} = \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{A}_*] \quad with \quad \mathbf{A}_* = (\lambda\mathbf{K}^\top\mathbf{K})^{-1}. \tag{18}$$

The problem with the distribution in (18) is that the prior 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 conjugate multivariate Bayesian analysis, as long as the likelihood function is normally distributed with full rank covariance matrix: Then 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. In the smoothing case we have to specify a large λ parameter, and 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 smoothing model to the transformed (= differenced) form of the model. This means that the parameter $\boldsymbol{\tau}$ is smoothed in the Bayesian model towards a function that minimizes the second order difference of the $\boldsymbol{\tau}$'s.

Now we can follow the recommendation of a $\lambda = 1600$ from a Bayesian point of view. If the series to be smoothed is given in quarterly growth rates, a standard deviation of $\sigma = 5\%$ seems to be reasonable. Now we have to come up with a guess of how big the variance of a smoothed series could or should be. The proposal of Hodrick and Prescott (1997, p.4) was: not more than an eighth of a percent or $\sigma_\tau^2 = 1/8$. This leads to the hypothetical sample size

$$\lambda = \sigma^2/\sigma_\tau^2 = 5^2/(1/8)^2 = 25 * 64 = 1600 \tag{19}$$

and demonstrates clearly the subjectivity of the assumption "smooth". (For $\sigma = 4\%$ we get $\lambda = 4^2/(1/8)/^2 = 32^2 = 1024$, for $\sigma = 6\%$ we get $\lambda = 6^2/(1/8)/^2 = 48^2 = 2304$.) From Table 1 we see that the residual standard deviation after removing the linear trend is about 6 per cent. As in many cases subjective priors can be justified by ex-post rationalization: If the result is smooth enough, like e.g. a thick line, then the (prior) assumptions are acceptable. In other words, to produce a smooth trend in this regression model,

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

[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 $\mathbf{A}_*^{-1}$.

we have to add 1600 hypothetical observations that the prior mean of $\boldsymbol{\tau}$ is zero.

It is interesting to note that both, the classical HP and the Bayesian smoothing requires strong prior information. In Bayesian terms this is made explicit through the assumption of a prior distribution, while in classical terms this information is implicitly hidden in the term "smoothing parameter". But using strong priors require special justification since it does not follow the 'principle of objectivity' or 'non-involvement of non-data information' that is so often promoted in classical inference for regression coefficients. Thus we are confronted with 2 types of parameters: the trend (nuisance) parameter $\boldsymbol{\tau}$ and the focus parameter $\boldsymbol{\beta}$ of the regression model. For the inference of $\boldsymbol{\beta}$ we try to minimize the influence of the prior (and choose small n'), while for the smoothing problem we estimate $\boldsymbol{\tau}$ and 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 7. In the HP smoothing model this is a matrix weighted average between the prior location $\mathbf{0}$ and the ML location $\mathbf{y}$. Note that in the Bayesian framework it does not matter that the $\boldsymbol{\tau}$ parameter has T components, i.e. as many parameters as there are observations, as long as there is a proper prior distribution.

3.1. A conjugate multi-normal-gamma (mNG) model for HP smoothing

First, we describe the conjugate smoothing approach that is in analogy to the Normal-Gamma sampling (NGS) model that can be found e.g. in Polasek (2010).

We consider the conjugate multi-normal-gamma (mNG) model for the inference of an unknown mean $\boldsymbol{\tau}$ in a univariate sampling problem (with sample size n) as in 16:

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_n], \quad or \quad \mathbf{y} \sim \mathcal{N}[\boldsymbol{\tau}, \sigma^2 \boldsymbol{\Sigma}_0],$$

where $\boldsymbol{\Sigma}_0$ denotes a known covariance matrix. 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 3. Thus we show, that the Bayesian simple HP smoothing model follows a multivariate version of the normal-gamma sampling (NGS) inference scheme with a highly informative and well defined prior covariance matrix $\mathbf{A}_*$.

Theorem 3 (The multivariate normal-gamma sampling (mNGS) model).
We consider the smoothing model in (16), then the conjugate Bayesian inference with $\mathbf{A}_* = (n'\mathbf{K}^\top \mathbf{K})^{-1}$

in the prior density as in (18) can be done in the following way.

> *The prior distribution is given as a normal-gamma density*
>
> $$(\boldsymbol{\tau}, \sigma^{-2}) \sim \mathcal{N}_n\Gamma[\boldsymbol{\tau}_*, \mathbf{A}_*, s_*^2, n_*]$$
>
> *and the likelihood of the observed data*
>
> $$\mathcal{Y} = \{\mathbf{y}_i \sim \mathcal{N}[\boldsymbol{\tau}, \sigma^2\boldsymbol{\Sigma}_0], \quad i = 1, \dots, n\}$$
>
> *yields the posterior distribution*
>
> $$(\boldsymbol{\tau}, \sigma^{-2}) \mid \mathcal{Y} \sim \mathcal{N}_n\Gamma[\boldsymbol{\tau}_{**}, \mathbf{A}_{**}, s_{**}^2, n_{**}].$$

with the parameters

$$\begin{aligned}
\boldsymbol{\tau}_{**} &= \mathbf{A}_{**}(n'\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau}_* + \boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{y}}),\\
\mathbf{A}_{**}^{-1} &= n'\mathbf{K}^\top\mathbf{K} + \boldsymbol{\Sigma}_0^{-1},\\
n_{**} &= n_* + n,\\
n_{**}s_{**}^2 &= n_*s_*^2 + \mathbf{y}^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K} + \boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0\mathbf{y}.
\end{aligned} \tag{20}$$

The error sum of squares (ESS) is $ns^2 = (\mathbf{y} - \hat{\boldsymbol{\tau}})^\top\boldsymbol{\Sigma}_0^{-1}(\mathbf{y} - \hat{\boldsymbol{\tau}}) = 0$ *as the OLS estimator in the homolog regression model is* $\hat{\boldsymbol{\tau}} = \mathbf{y}$ *and* α *is the discrepancy term that serves as a penalty term for the variance in all conjugate models.*

Proof 3.

The likelihood of the above smoothing model (16) is simply derived from $\mathbf{y} \sim \mathcal{N}[\boldsymbol{\tau}, \sigma^2\boldsymbol{\Sigma}_0]$*. Let us define a 'multi-normal-gamma' prior, leading to a family of mNG conjugate distribution that follows as a multivariate extension from the normal-gamma (NG) distribution:*

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

where $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$ *is a known covariance matrix.*[6] *Similar as for the* $N\Gamma$ *distribution we define the* $mN\Gamma$ *distribution as*

$$\begin{aligned}
p(\boldsymbol{\tau}, \sigma^{-2}) &= p(\boldsymbol{\tau} \mid \sigma^{-2})p(\sigma^{-2}) = \mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_*, \sigma^2(n'\mathbf{K}^\top\mathbf{K})^{-1}]\ \Gamma[\sigma^{-2} \mid s_*^2, n_*]\\
&\propto exp\left\{-\frac{1}{2\sigma^2}\left((\boldsymbol{\tau} - \boldsymbol{\tau}_*)^\top n'\mathbf{K}^\top\mathbf{K}(\boldsymbol{\tau} - \boldsymbol{\tau}_*)\right)\right\} exp\left\{-\frac{1}{2\sigma^2}n_*s_*^2\right\}.
\end{aligned} \tag{21}$$

Therefore the joint prior of the $mNG = \mathcal{N}_n\Gamma$ *distribution has the form*

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

This has the structure of a univariate normal-gamma ($N\Gamma$ *) distribution but now the* $\boldsymbol{\tau}$ *vector is n-dimensional. We find the posterior mNG distribution by multiplying the prior with the likelihood:*

---

[6](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.)

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

We have to apply the theorem of combining the 2 quadratic forms in $\boldsymbol{\tau}$ (see (7) in the Appendix section 6) to get

$$
(\boldsymbol{\tau}-\boldsymbol{\tau}_{**})^\top\mathbf{A}_{**}^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}_{**})+(\mathbf{y}-\boldsymbol{\tau}_*)^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K}+\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0(\mathbf{y}-\boldsymbol{\tau}_*), \qquad (23)
$$

and the parameters $\boldsymbol{\tau}_{**}$ and $\mathbf{A}_{**}$ are given as in (20). The second term in (23) is called discrepancy term between the observation $\mathbf{y}$ and the prior location $\boldsymbol{\tau}_*$ which is in the HP smoothing model zero. Thus the discrepancy term is for $\boldsymbol{\tau}_*=\mathbf{0}$ reduces to

$$
\begin{aligned}
\alpha &= \mathbf{y}^\top n'\mathbf{K}^\top\mathbf{K}(n'\mathbf{K}^\top\mathbf{K}+\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0\mathbf{y} = \\
&= \mathbf{y}^\top(\mathbf{C}^\top\mathbf{C}/n'+\boldsymbol{\Sigma}_0^{-1})^{-1}\mathbf{y}=\mathbf{y}^\top\mathbf{Q}\mathbf{y} \qquad (24)
\end{aligned}
$$

Furthermore, we can write the posterior sum of squares matrix

$$
n_{**}s_{**}^2 = n_*s_*^2+\tilde{\alpha}_i. \qquad (25)
$$

The quadratic form of the posterior multi-normal-gamma density $\mathcal{N}_n\Gamma$ for $p(\boldsymbol{\tau},\sigma^{-2})$ in (23) can be factored into a conditional normal times a gamma distribution

$$
p(\boldsymbol{\tau} \mid \sigma^{-2})\; p(\sigma^{-2}) = \mathcal{N}_n[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \sigma^2\mathbf{A}_{**}]\; \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}] \qquad (26)
$$

where the marginal distribution for $\boldsymbol{\tau}$ is a multivariate t distribution with n_{**} d.f. given by

$$
\boldsymbol{\tau} \mid \mathbf{y} \sim \mathrm{t}_T\left[\boldsymbol{\tau}_{**}, s_{**}^2\mathbf{A}_{**}, n_{**}\right]. \qquad (27)
$$

In the Bayesian case, the smoothness predictor of the observations in $\mathbf{y}$ is given by the posterior distribution of $\boldsymbol{\tau}$. The point estimate of the smoother is the point estimate of the posterior distribution. A common choice is the posterior mean which is given by (20)

$$
\boldsymbol{\tau}_{**} = \mathbf{A}_{**}(n'\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau}_* + \boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{y}}), \qquad (28)
$$

where $\bar{\mathbf{y}}$ is the mean of the sample. For one observation $\bar{\mathbf{y}}=\mathbf{y}$ and homoskedastic errors $\boldsymbol{\Sigma}_0=\mathbf{I}_n$ the posterior mean gives the same formula as in the HP smoother in the classical case in (15): $\widehat{\mathbf{y}} = \boldsymbol{\tau}_{**}$ and

$$\boldsymbol{\tau}_{**} = (\mathbf{I}_T + \lambda \mathbf{K}^\top \mathbf{K})^{-1}\mathbf{y} = (\mathbf{I}_T - \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K})\mathbf{y}. \tag{29}$$

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 in Bayesian analysis by the predictive density, where the point prediction is obtained again via the posterior mean as in (28).

3.2. The predictive density for τ_{T+1} in the conjugate HP model

If we want to make a prediction in the HP smoothing model with homolog parameters $\boldsymbol{\tau}$, we are confronted with the situation that for period $T+1$ we can generate the next smooth from the previous estimated parameters stemming from the restriction $\tau_{T+1} - 2\tau_T + \tau_{T-1} = 0$ leading to

$$\tau_{T+1} \mid \boldsymbol{\tau}, \sigma^2 \sim \mathcal{N}[2\tau_T - \tau_{T-1}, \sigma^2], \tag{30}$$

where σ^2 is the residual variance of the HP model. Given the posterior distribution for $\boldsymbol{\tau}$ we get as point prediction for the smooth at

$$\widehat{\tau}_{T+1} = 2\tau_T^{**} - \tau_{T-1}^{**}. \tag{31}$$

We can write this one-step-ahead forecast (30) as a linear combination

$$\tau_{T+1} = \mathbf{q}^\top \boldsymbol{\tau} \quad \textit{with} \quad \mathbf{q}^\top = (0, ..., 0, -1, 2) : T \times 1.$$

From Theorem 3 the posterior distribution for the HP smoother $\boldsymbol{\tau}$ is given by $\mathcal{N}_T\Gamma[\boldsymbol{\tau}_{**}, \mathbf{A}_{**}, s^2_{**}, n_{**}]$ and assuming $\boldsymbol{\Sigma}_0 = \mathbf{I}_T$ the parameters, which are given in (20) simplify to

$$\boldsymbol{\tau}_{**} = \mathbf{A}_{**}(n'\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau}_* + \mathbf{y}), \quad \textit{and} \quad \mathbf{A}_{**}^{-1} = n'\mathbf{K}^\top\mathbf{K} + \mathbf{I}_T.$$

Therefore the normal part of the posterior NG distribution leads via the recursion formula (31) to the conditional predictive density

$$\tau_{T+1} \mid \sigma^2 \sim \mathcal{N}[\mathbf{q}^\top\boldsymbol{\tau}_{**}, \sigma^2_{T+1} = \sigma^2\mathbf{q}^\top\mathbf{A}_{**}\mathbf{q}]. \tag{32}$$

Integrating over σ^{-2} in this linear combination of a NG distribution gives an unconditional t distribution with n_{**} df:

$$\tau_{T+1} \sim t[\mathbf{q}^\top \boldsymbol{\tau}_{**}, \sigma^2_{**}\mathbf{q}^\top \mathbf{A}_{**}\mathbf{q}, n_{**}]. \tag{33}$$

For the prediction of the next two parameters in $\boldsymbol{\tau}_f = (\tau_{T+1}, \tau_{T+2})^\top$ we have to define the matrix $\mathbf{Q}^\top = \binom{\mathbf{q}^\top}{\mathbf{e}_T^\top} = \begin{pmatrix} 0 & \dots & 0 & -1 & 2 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix}$: $T \times 2$ where $\mathbf{e}_T$ is the T-th unity vector. Thus, we can predict the next 2 future y observation by the conditional distribution of the Bayesian HP smoothing model according to (16)

$$\mathbf{y}_f \mid \sigma^2 \sim \mathcal{N}_2[\boldsymbol{\tau}_f, \sigma^2 \mathbf{I}_2] \quad with \quad \boldsymbol{\tau}_f = \mathbf{Q}\boldsymbol{\tau}. \tag{34}$$

The distribution of the future smoothing parameters $\boldsymbol{\tau}_f = \mathbf{Q}\boldsymbol{\tau}$ can be derived from the posterior density (28) using the matrix $\mathbf{Q}$: This is the conditional predictive density (34) for $\boldsymbol{\tau}_f$

$$\boldsymbol{\tau}_f \mid \sigma^2 \sim \mathcal{N}_2[\boldsymbol{\tau}_{f**}, \sigma^2 \mathbf{Q}\mathbf{A}_{**}\mathbf{Q}^\top] \quad with \quad \boldsymbol{\tau}_{f**} = \mathbf{Q}^\top \boldsymbol{\tau}_{**}. \tag{35}$$

3.3. The predictive density for the mNG distribution

The multi-normal-gamma distribution is defined as a conditional normal distribution times a gamma distribution. Therefore we can derive the predictive distribution in the usual way of a conjugate normal-gamma model. Let $\mathbf{y}_f = (y_{T+1}, y_{T+2})^\top$ be the two observation we want to predict into the future to avoid the end point "turbulence" created by the smoothness prior of the HP approach. Prediction of the next 2 observations in the HP smoothing model is

$$\begin{pmatrix} y_{T+1} \\ y_{T+2} \end{pmatrix} = \begin{pmatrix} \tau_{T+1} \\ \tau_{T+2} \end{pmatrix} + \begin{pmatrix} u_{T+1} \\ u_{T+2} \end{pmatrix} \tag{36}$$

and in matrix form this smoothing model for the future data set, indexed by f, is $\mathbf{y}_f = \boldsymbol{\tau}_f + \mathbf{u}_f$ or

$$\mathbf{y}_f \sim \mathcal{N}_2[\mathbf{Q}\boldsymbol{\tau}, \sigma^2 \mathbf{I}_2] = p(\mathbf{y}_f \mid \boldsymbol{\theta}), \tag{37}$$

because the parameter in the normal-gamma model are $\boldsymbol{\theta} = (\boldsymbol{\tau}, \sigma^{-2})$ we obtain the posterior predictive distribution for the (simple) Bayesian HP model in the following way

$$p(\mathbf{y}_f \mid \mathbf{y}) = \int \int p(\mathbf{y}_f \mid \boldsymbol{\tau}, \sigma^{-2})\ p(\boldsymbol{\tau}, \sigma^{-2} \mid \mathbf{y})\ d\boldsymbol{\tau}\ d\sigma^{-2}. \tag{38}$$

Integration is done via the posterior normal-gamma density, given by (26) as

$$(\boldsymbol{\tau}, \sigma^{-2} \mid \mathbf{y}) = \mathcal{N}_n\Gamma[\boldsymbol{\tau}_{**}, \mathbf{A}_{**}, s^2_{**}, n_{**}]$$

and the conditional predictive density is given by (34) as $p(\mathbf{y}_f \mid \boldsymbol{\tau}, \sigma^{-2}) = \mathcal{N}_2[\mathbf{y}_f \mid \boldsymbol{\tau}_{f**} = \mathbf{Q}\boldsymbol{\tau}_{**}, \sigma^2\boldsymbol{\Sigma}_f]$ with $\boldsymbol{\Sigma}_f = \mathbf{Q}\mathbf{A}_{**}\mathbf{Q}^\top$ as in (35).

The joint predictive distribution of $\mathbf{y}_f$ and the parameter $\boldsymbol{\theta} = (\boldsymbol{\tau}, \sigma^{-2})$ is given by

$$\begin{aligned} p(\mathbf{y}_f, \boldsymbol{\theta} \mid \mathbf{y}) &= p(\mathbf{y}_f \mid \boldsymbol{\theta}, \mathbf{y}) \cdot p(\boldsymbol{\theta} \mid \mathbf{y}) \\ &= \mathcal{N}[\mathbf{y}_f \mid \mathbf{Q}\boldsymbol{\tau}, \sigma^2\mathbf{I}_2]\ \mathcal{N}\left[\boldsymbol{\tau} \mid \mathbf{Q}\boldsymbol{\tau}_{**}, \mathbf{A}_{**}\right]\ \Gamma[\sigma^{-2} | s^2_{**}, n_{**}] \\ &= \mathcal{N}\Gamma\left[\tau_{f**}, \boldsymbol{\Sigma}_f, s^2_{**}, n_{**}\right] \end{aligned}$$

with the parameters as in (35)). Next, we derive the predictive distribution for the multi-normal-gamma sampling (mNGS) model.

Theorem 4 (Prediction in the multi-normal-gamma sampling (mNGS) model). *We consider the conditional prediction problem (34) of the Bayesian HP smoothing model to predict the observations at time T and $T-1$.*

The posterior distribution

$$p(\boldsymbol{\tau}, \sigma^{-2} \mid \mathbf{y}) = \mathcal{N}_T\Gamma[\boldsymbol{\tau}_{**}, \mathbf{A}_{**}, s^2_{**}, n_{**}]$$

and the conditional predictive distribution for $\mathbf{y}_f = \mathbf{Q}^\top\boldsymbol{\tau}$, given by

$$\mathbf{y}_f \mid \boldsymbol{\tau}, \sigma^2 \sim \mathcal{N}_2[\boldsymbol{\tau}_{f**}, \sigma^2\boldsymbol{\Sigma}_f]$$

yield the (2-dimensional) predictive distribution

$$p(\mathbf{y}_f \mid \mathbf{y}) = \mathrm{t}_2\left[\boldsymbol{\tau}_{f**}, \boldsymbol{\Sigma}_f s^2_{**}, n_{**}\right].$$

with $\boldsymbol{\tau}_{f**} = \mathbf{Q}^\top\boldsymbol{\tau}_{**}$ and $\boldsymbol{\Sigma}_f = \mathbf{Q}^\top\mathbf{A}_{**}\mathbf{Q}$.

Proof 4. *Using the formulas of the* t*and gamma integrals we find for the (marginal) posterior predictive distribution for $\mathbf{y}_f$*

$$\begin{aligned} p(\mathbf{y}_f \mid \mathbf{y}) &= \int \mathcal{N}\left[\boldsymbol{\tau}_{f**}, \sigma^2\boldsymbol{\Sigma}_f\right] \cdot \Gamma[\sigma^{-2} \mid s^2_{**}, n_{**}]\, d\sigma^{-2} \\ &= \mathrm{t}_2\left[\boldsymbol{\tau}_{f**}, \boldsymbol{\Sigma}_f s^2_{**}, n_{**}\right], \end{aligned} \tag{39}$$

a bivariate t *distribution with mean $\boldsymbol{\tau}_{f**} = \mathbf{Q}^\top\boldsymbol{\tau}_{**}$ and n_{**} df.*

3.4. Improved HP smoothers using endpoint predictions

The previous prediction of the $T+1$-st smoother can be used for an improved HP smoother. Using the point predictor we define the augmented observation vector $\tilde{\mathbf{y}}^\top = (\mathbf{y}^\top, \mathbf{y}_f^\top)$ and the differencing matrix $\tilde{\mathbf{K}} : T \times T+2$ is given by

$$\tilde{\mathbf{K}} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & ... & 0 & 0 & 0 \\ & ... & & ... & ... & ... & ... & ... & \\ 0 & 0 & 0 & 0 & 0 & ... & 1 & -2 & 1 \end{pmatrix}. \tag{40}$$

The improved HP smoother is now given by the augmented $\tilde{\mathbf{K}}$ matrix and the augmented observations $\tilde{\mathbf{y}} = \binom{\mathbf{y}}{\mathbf{y}_f}$:

$$\tilde{\boldsymbol{\tau}}_{**} = [\mathbf{I}_{T+2} + \lambda \tilde{\mathbf{K}}^\top \tilde{\mathbf{K}}]^{-1} \tilde{\mathbf{y}} = \tilde{\mathbf{A}}_{**} \tilde{\mathbf{y}}. \tag{41}$$

For the improved HP smoother we use only the first T components in $\tilde{\mathbf{y}}_{**} : \ddot{\mathbf{y}}_{HP}$.

4. Model selection and Bayes testing

In this section we show how to compute the Bayesfactor for the HP smoother and to select the order of smoothness prior by marginal likelihoods. The assumption to do this is a normal prior distribution with full rank, Therefore we augment the differencing matrix by the T-th unity vector in order to get an invertible prior covariance matrix. The first order differencing matrix is $\mathbf{K}_1$ -nd the higher order differencing matrices are matrix powers: $\mathbf{K}_i = \mathbf{K}_1^i$ and therefore the prior covariance matrix of the i-th smoothness model is $\mathbf{A}_{i*} = (\lambda \mathbf{K}_i^\top \mathbf{K}_i)^{-1} = (\mathbf{K}^\top \mathbf{K})^{-i}/\lambda$. For the conjugate normal-gamma regression model the marginal likelihood can be computed in closed form as the next theorem shows.

Theorem 5 (The marginal likelihood for the Bayesian HP model).

The marginal (data) likelihood (MDL) of the HP regression model is given by a product of 3 factors (that are 3 ratios of prior to posterior parameters):

$$p(\mathbf{y} \mid HP) = (\pi)^{-\frac{n}{2}} \frac{|\mathbf{A}_{**}|^{\frac{1}{2}}}{|\mathbf{A}_*|^{\frac{1}{2}}} \times \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_*}{2})} \times \frac{(n_* s_*^2/2)^{\frac{n_*}{2}}}{(n_{**} s_{**}^2/2)^{\frac{n_{**}}{2}}}, \tag{42}$$

where n_{**} and s_{**}^2 are the posterior parameter given in (20) of Theorem 3.

Note that the marginal data likelihood for the HP model follows the ordinary MDL formula for the normal-gamma sampling model $MDL_{HP} = p_{HP}(\mathbf{y})$

$$p_{HP}(\mathbf{y}) = (\pi)^{-\frac{n}{2}} \times R_{det} \times R_{df} \times R_{ESS}$$

the ratio of determinants (R_{det}), *the ratio of d.f.* (R_{df}), *and the ratio of residual variances* (R_{ESS}). *Usually it is better to compute the* $lml = log(MDL)$ *given by*

$$lml_{HP} = -\frac{n}{2}log(\pi) + log(R_{det}) + log(R_{df}) + log(R_{ESS}). \quad (43)$$

The lml_{HP} *times* -2 *is (with* $ESS_* = n_* s_*^2/2$ *and* $ESS_{**} = n_{**} s_{**}^2/2$*)*

$$-2lml_{HP} = nlog(\pi) + log\frac{|\,\mathbf{A}_{**}^{-1}\,|}{|\,\mathbf{A}_{*}^{-1}\,|} + (n_{**}-3)log(2) + 2log(\Gamma(n_{**}-1)) + n_* log(ESS_*) - n_{**} log(ESS_{**}). \quad (44)$$

The ratio of determinants in the HP model is computed by the inverses

$$R_{det} = |\,\mathbf{A}_*\,|\,|\,\mathbf{A}_{**}^{-1}\,| = |\,\lambda\mathbf{K}^\top\mathbf{K}(\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K})^{-1}\,| = |\,(\mathbf{I}_T + (\lambda\mathbf{K}^\top\mathbf{K})^{-1})^{-1}\,|$$

with $\mathbf{A}_* = (\lambda\mathbf{K}^\top\mathbf{K})^{-1}$ *and* $\mathbf{A}_{**} = (\mathbf{I}_T + \lambda\mathbf{K}^\top\mathbf{K})^{-1}$ *where* $\mathbf{K}$ *is the tridiagonal differencing matrix.*

Proof 5. *The ratio of determinants for differencing matrices of order* i *is*

$$R_{det,i}^2 = \frac{|\,\mathbf{A}_{i*}^{-1}\,|}{|\,\mathbf{A}_{i**}^{-1}\,|} = \frac{|\,\lambda\mathbf{K}_i^\top\mathbf{K}_i\,|}{|\,\mathbf{I}_n + \lambda\mathbf{K}_i^\top\mathbf{K}_i\,|}$$

and the ESS ratio can be computed as

$$R_{ESS} = \frac{(n_* s_*^2/2)^{\frac{n_*}{2}}}{(n_{**} s_{**}^2/2)^{\frac{n_{**}}{2}}} = \frac{(n_* s_*^2/2)^{\frac{n_*}{2}}}{((n_* s_*^2 + \alpha_i)/2)^{\frac{n_{**}}{2}}}.$$

The ratio of d.f. (R_{df}) *is given for* $n_* = 1$ *by* $\Gamma(1/2) = \sqrt{\pi}$ *and therefore we find*

$$R_{df} = \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_*}{2})} = \frac{(n_{**}-2)!!}{2^{(n_{**}-1)/2}} \quad (45)$$

The log df-ratio is

$$\begin{aligned} log(R_{df}) &= log\frac{(n_{**}-2)!!}{2^{(n_{**}-1)/2}} = (n_{**}-2)log(2) + log((n_{**}-2)!) - \frac{(n_{**}-1)}{2}log(2) = \\ &= \frac{(n_{**}-3)}{2}log(2) + log(\Gamma(n_{**}-1)), \quad (46) \end{aligned}$$

because $log(n_{**}-2)!! = (n_{**}-2)log(2) + log((n_{**}-2)!)$ *and the double factorial is defined as* $(2k)!! = 2^k \cdot k!$.

Theorem 6 (Bayes test between HP models of different smoothness order).

For the Bayes test between two HP models of order i and j we need the Bayes factor (BF), which is defined as the ratio of the 2 marginal likelihoods of HP models and the BF is given by

$$BF = \frac{p(\mathbf{y} \mid HP_i)}{p(\mathbf{y} \mid HP_j)} = \frac{R_{ESS,i}}{R_{ESS,j}} \frac{R_{det,i}}{R_{det,j}}$$

and the log BF is computed by

$$log(BF) = \frac{n_{**}}{2} log\left(n_* s_*^2 + \alpha_i\right) + \frac{p}{2} log(n_2)$$

with α_i the discrepancy factor of the smoothness model of order i:

$$\alpha_i = \mathbf{y}^\top n' \mathbf{K}_i^\top \mathbf{K}_i (n' \mathbf{K}_i^\top \mathbf{K}_i + \mathbf{\Sigma}_0)^{-1} \mathbf{\Sigma}_0 \mathbf{y} \tag{47}$$

Note that if $n_2 = 1$ then $log(n_2) = 0$ and the second term vanishes.

Proof 6. *The BF is given by the ratio of marginal likelihoods, with the ESS ratio of ratios (RoR) given by*

$$\frac{R_{ESS,i}}{R_{ESS,j}} = \frac{((n_* s_*^2 + \alpha_i)/2)^{\frac{n_{**}}{2}}}{((n_* s_*^2 + \alpha_j)/2)^{\frac{n_{**}}{2}}} = \left(\frac{n_* s_*^2 + \alpha_i}{n_* s_*^2 + \alpha_j}\right)^{\frac{n_{**}}{2}}$$

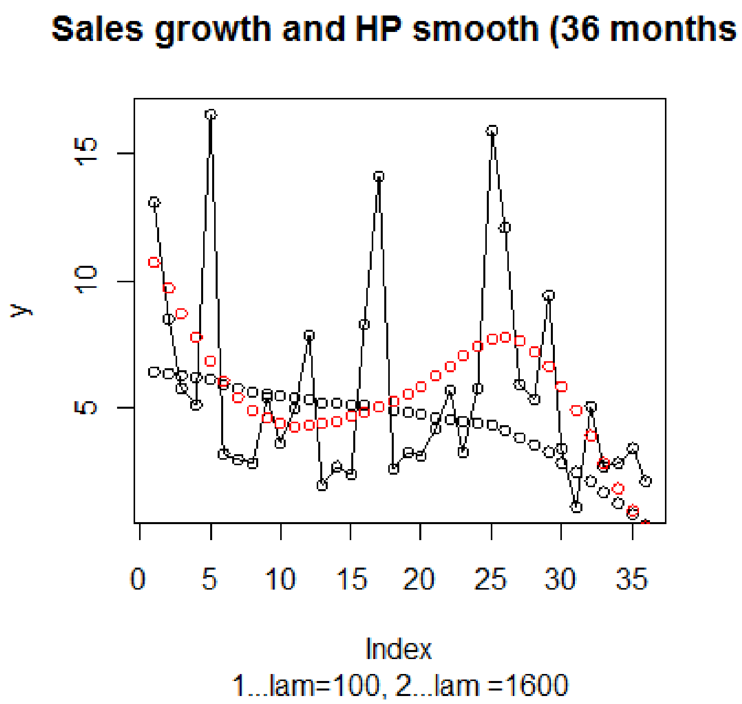
and the determinant ratio given by

$$\frac{R_{det,i}^2}{R_{det,j}^2} = \frac{|\lambda \mathbf{K}_i^\top \mathbf{K}_i|}{|\mathbf{I}_n + \lambda \mathbf{K}_i^\top \mathbf{K}_i|} / \frac{|\lambda \mathbf{K}_j^\top \mathbf{K}_j|}{|\mathbf{I}_n + \lambda \mathbf{K}_j^\top \mathbf{K}_j|}$$

because the R_{df} and the constant involving π cancels out.

Example for the Bayes test: We compare several HP models by log marginal likelihoods and we look for the best smoothing order (see the R program in the appendix). We have observed a time series of sales growth over 36 month and we are interested in the long-term trend. Figure 1 panel (a) shows the data with the HP smooth of order 1 and 2. For order 1 we used the smoothing constant $\lambda = 100$, because of relaxing the smoothness variance in (19) to 1/2, i.e. $\lambda = \sigma^2/\sigma_\tau^2 = 5^2/(1/2)^2 = 25 * 4 = 100$ or equivalently by $10^2/1$, while for order 2 the usual $\lambda = 1600$. The reduction of 16 was needed in order to accommodate the larger variance of the first differences. We see the falling trend that aggravates in the last 12 months. In panel (b) we have plotted the lml for smoothness order 1 up to 5 and we see that the maximum is attained at order 2.

The associated Bayes factors for comparing the sequence of Bayes tests is given by: $BF_1^2 = 65018.05, BF_2^3 = 0.0046, BF_1^2 = 0.0004, BF_1^2 = 0.0003$. Thus, there is overwhelming evidence that order 2 smoothness is the best choice for HP smoothing.



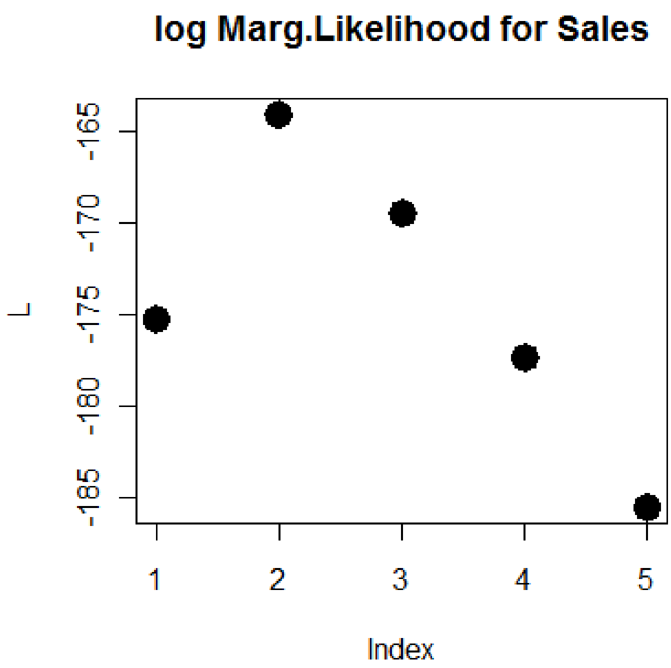


Figure 1: a) HP smooth of sales index & data b) log ML picks k=2 as HP smoothing order

5. Summary

This paper has shown that the HP filter is an over-parameterized regression problem from a Bayesian point that can be estimated in a conjugate normal-gamma model with a strong prior on the smoothness component. The large value of the smoothness parameter λ serves in the Bayesian model as a hypothetical sample size for the value of the prior information. To produce a smooth output one has to increase the prior precision to stick quite close to the chosen "smoothness" prior, which is defined by the second difference of the smooth component, i.e. the parameter vector to be estimated.

Furthermore, the HP filter in a conjugate regression model allows to derive the predictive distribution of the HP smooth. This is important addition if we like to complete the HP smooth on both ends of the time series, if we use non-invertible differencing matrices. HP smoothing can be simplified if we use invertible differencing matrices as it is shown in Polasek (2011).

The proposed Bayesian view of the HP procedure opens a new modeling technique for smoothing output variables in more complex econometric models. These are models that require more adjustments and simplifications before the smoothing can be done. The Bayesian interpretation of HP models shows how to obtain more flexibility via the prior information that is used for the estimation of the smooth and the non-smooth part in such a complex smoothing models. The non-conjugate estimation of the HP model uses the MCMC approach and allows application of the HP smoothing approach to extended HP models for non.stationary data and to a spatial smoothing model as it is discussed in Polasek (2011).

6. Appendix: Results on Combination of Quadratic Forms

We list the standard result for combining quadratic forms in normal Bayes models and the associated MESS decomposition:

Theorem 7 (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*

$$\begin{aligned}&(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{H}(\boldsymbol{\beta}-\mathbf{b})+(\boldsymbol{\beta}-\mathbf{b}_*)^\top\mathbf{H}_*(\boldsymbol{\beta}-\mathbf{b}_*)\\ &=(\boldsymbol{\beta}-\mathbf{b}_{**})^\top\mathbf{H}_{**}(\boldsymbol{\beta}-\mathbf{b}_{**})+(\mathbf{b}-\mathbf{b}_*)^\top\mathbf{H}_*(\mathbf{H}_*+\mathbf{H})^{-1}\mathbf{H}(\mathbf{b}-\mathbf{b}_*)\end{aligned} \tag{48}$$

with the parameters

$$\begin{aligned}\mathbf{H}_{**} &= \mathbf{H}_*+\mathbf{H},\\ \mathbf{b}_{**} &= \mathbf{H}_{**}^{-1}(\mathbf{H}_*\mathbf{b}_*+\mathbf{H}\mathbf{b}).\end{aligned} \tag{49}$$

7. R program for log marginal likelihoods

```
lmlHP=function(y,nx,sx,k,lam=1600) {
# fct for logML in NGR model  #WP Aug11
#for square invertible diff. matrices K
#bx     #prior (post)mean beta
#Ax (Axx)   #prior (post) cov beta
#sx (sxx) #prior (post) sigma^2
#nx (nxx) #prior (post)df of sig
#k  --- differencing order
n=length(y) #nobs
eye = diag(n) #id matrix
K1=diff(diag(n),lag=1,d=1);
K1=rbind(K1,eye[n,])        #make it square
K=K1              # n x n diff matrix
if (k>1) for(i in c(1: k)) K= t(K)%*%K1 #recursion
nxx = nx + n             #post df ... sig
Axi=lam*t(K)%*%K     #prior 4 smooth
Axx=solve(Axi + eye)      #post 4 smooth
alph= t(y)%*%Axi%*%Axx%*%y  #discrepancy
sxx=(nx*sx + alph)/nxx     #post sigma^2
r1= det(Axx) * det(Axi)     #det ratio
r2=lgamma(nxx/2) -lgamma(nx/2)            #log n ratio
r3=nx/2*log(sx*nx/2)-nxx/2*log(sxx*nxx/2) #log ESS
m=-.5*pi + .5 *log(r1) + r2 + r3        #logML sum
return(m)   #out = log ML
}
 #sales: compute up to order 5 the lml:
L=rep(0,5); for(i in c(1: 5)) L[i]=  lmlHP(y,nx,sx=1,i)
```

```
round(exp(diff(L)),4)       #Bayes factors
##############################################
HPfilter = function(x,k,lambda=1600){
#Function for simple HP smoothing up to order 5; #WP sept 2011
n=length(x)
eye <- diag(length(x))
K1=-diff(diag(n),lag=1,d=1); K1=rbind(K1,eye[n,])
K=K1  # n x n diff matrix  det(K) dim(K)
if (k>1) for(i in c(1: k)) K= t(K)%*%K1
result <- solve(eye+lambda*t(K)%*%K)%*%x
return(result)
}
HPfilter(y36,2)
```

8. References

Baxter, M. and R. 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, K. and M. O. Ravn (1991), Contemporary Macroeconomic Fluctuations: an International Perspective, Discussion Paper 1991-06, University of Southampton.

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

Cooley, T. F. and Prescott, E. C. (1995), Economic growth and business cycles. In Cooley, T. F., editor, Frontiers of Business Cycle Research, chapter 1. Princeton University Press, Princeton.

Da Fonseca C.M. (2005), On the eigenvalues of some tridiagonal matrices, J. Comput. Appl. Math. 200 (2007), no. 1, 283-286. Departamento de Matematica, Universidade de Coimbra, Preprint 05-16. http://www.mat.uc.pt/preprints/ps/p0516.pdf

de Jong P. and N. Shepard (1995), The simulation smoother for time series models, Biometrika 8 2 , 2, pp. 339-50

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.

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

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

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.

Leser C.E.V. (1961), A Simple Method of Trend Construction, Journal of the Royal Statistical Society B, 23, 91-107.

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

Polasek W. (2010), Bayesian econometrics with R, IHS Vienna, lecture notes.

Polasek W. (2011a), MCMC estimation of extended Hodrick-Prescott (HP) filtering models, IHS-Vienna and RCEA discussion paper.

Polasek W. (2011b), The extended Hodrick-Prescott (eHP) filter as a conjugate regression model, IHS-Vienna (Institute for Advanced Studies) discussion paper.

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

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

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