The Rimini Centre for Economic Analysis

WP 11-45

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

THE EXTENDED HODRICK-PRESCOTT (HP) FILTER FOR SPATIAL REGRESSION SMOOTHING

The extended Hodrick-Prescott (HP) filter for spatial regression smoothing

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 long-term component of stationary series like growth rates. The new extended HP smoothing model is applied to data-sets with an underlying metric and requires a Bayesian linear regression model with a strong prior based on differencing matrices for the smoothness parameter and a weak prior for the regression part. We define a Bayesian spatial smoothing model with neighbors for each observation and we define a smoothness prior similar to the HP filter in time series. This opens a new approach to model-based smoothers for time series and spatial models based on MCMC. We apply it to the NUTS-2 regions of the European Union for regional GDP and GDP per capita, where the fixed effects are removed by an extended HP smoothing model.

Keywords: Hodrick-Prescott (HP) smoothers, smoothed square loss function, spatial smoothing, smoothness prior, Bayesian econometrics.

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

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

Preprint submitted to Elsevier *October 25, 2011*

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).

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. 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. Furthermore, in Section 4 I show that this approach enables us to define 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 applied in for spatial regional GNP data in Europe. The approaches is based on a distance concept in order to define spatial nearest neighbors (NN). A final section concludes. The appendix contains a result on combination of quadratic forms.

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).

Hodrick and Prescott (1997) 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.

Recently Polasek (2011a,b) has shown that the Bayesian modeling for HP smoothing can be also done using the conjugate concept of a multivariate normal-gamma (mNG) model. Furthermore, the Bayesian approach allows model selection and Bayes testing.

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 $\mathbf{y} = [y_1, ..., y_T]^\top$. 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 (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 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* $\boldsymbol{\tau}$*:*

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

and we apply Theorem 7 for combining quadratic forms 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.*

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 highly informative prior:*

$$\mathbf{y} \sim \mathcal{N}[\boldsymbol{\tau}, \sigma^2 \mathbf{I}_T] \quad \textit{with} \quad \mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}]. \tag{10}$$

Theorem 1 has led to the following 'signal + noise' decomposition of the data $\mathbf{y}$:

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

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

$$\mathbf{P}_\lambda = \mathbf{K}^\top(\mathbf{I}_{T-2}\lambda^{-1} + \mathbf{K}\mathbf{K}^\top)^{-1}\mathbf{K} = \mathbf{I}_T - \mathbf{A}_{**}^{-1} \tag{11}$$

estimates the rough or noise component of the HP smoothing model.

3. The HP filter as a Bayesian smoothness regression model

The Bayesian HP type smoothing model starts also from the HP type regression (or 'smooth + noise decomposition') model (1), $\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T]$, with the identity matrix as "regressors" and where $\boldsymbol{\tau} : T \times 1$ is the pera-parametric parameter vector containing the smooth and the error term $\boldsymbol{\varepsilon}$, which 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 for time series smoothing is the second order

[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 $\mathbf{K}_p = \mathbf{K}_1^p$.

differencing matrix $\mathbf{K} : (T-2) \times T$:

$$\mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{T-2}]. \tag{12}$$

For spatial smoothing we can define a smoothing matrix as in (19) that is $\mathbf{K} : T \times T$ and invertible. In this special case with prior mean $\mathbf{0}$ it is easy to see that the prior is equivalent to[2] the distributional smoothness assumption for $\boldsymbol{\tau}$

$$\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{A}_*] \quad \textit{with} \quad \mathbf{A}_* = (\lambda\mathbf{K}^\top\mathbf{K})^{-1}. \tag{13}$$

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 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 (or expected noise to signal ratio)

$$\lambda = \sigma^2/\sigma_\tau^2 = 5^2/(1/8)^2 = 25 * 64 = 1600 \tag{14}$$

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

In the spatial context we can use the same reasoning for the smoothness constant as for time series, if the data set to be smoothed consists of e.g. growth rates across regions. We could relax the assumption that

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

the smooth should be 1/8 to 1/4 or 1/2 of a percentage point. In that case the smoothing constant becomes smaller: $\lambda = 400$ or 100. In case the cross sectional variance σ^2 goes up, say to 10%, then λ increases to 1600 or 400. Thus, we can expect for more volatile cross-sections about the same large λ's as for less volatile time series. For (cross-sectional) data sets that are not growth rates, the scaling is not important since the scale factor cancels out in the λ defined as a ratio (14) and as long as we agree we the above reasoning of what we expect to be smooth.

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 maximum likelihood location $\mathbf{y}$. Note that in the Bayesian framework it does not matter that the $\boldsymbol{\tau}$ parameter with T components is pera-parametric, i.e. as many parameters as there are observations, as long as there is a proper prior distribution.

4. 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)

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

where $\mathbf{W}$ is a row-normalized weight matrix, $\mathbf{W}\mathbf{y}$ is the first order spatial lag of $\mathbf{y}$, and ρ is the spatial correlation coefficient (see Lesage and Pace 2009). Model (15) can be viewed as a SAR(1) model is equivalent to the transformed model

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

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

$$\mathbf{y} \sim \mathcal{N}[\rho\mathbf{W}\mathbf{y} + \boldsymbol{\tau}, \sigma^2\mathbf{I}_n] \quad or \quad \mathbf{y} \sim \mathcal{N}[\mathbf{R}^{-1}\boldsymbol{\tau}, \sigma^2(\mathbf{R}^\top\mathbf{R})^{-1}] \tag{16}$$

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

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. Similar to the HP filter (3) for time series we can find the spatial HP smoother as the minimizer of the SSL function as in Definition 1

$$\begin{aligned} \boldsymbol{\tau}_{**} &= \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad with \\ SSL(\boldsymbol{\tau}) &= (\mathbf{R}\mathbf{y} - \boldsymbol{\tau})^\top(\mathbf{R}\mathbf{y} - \boldsymbol{\tau}) + \lambda \sum_{i=1}^{n} (\mathbf{w}_i^{(0)}\mathbf{y} - 2\mathbf{w}_i^{(1)}\mathbf{y} - \mathbf{w}_i^{(2)}\mathbf{y})^2. \end{aligned} \tag{17}$$

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

$$smooth = \sum_{i=1}^{n} (y_i - \mathbf{w}_i^{(1)}\mathbf{y} - \mathbf{w}_i^{(1)}\mathbf{y} + \mathbf{w}_i^{(2)}\mathbf{y})^2 = \sum_{i=1}^{n} (\Delta^{(2)}\mathbf{w}_i\mathbf{y})^2 = \mathbf{y}^\top\mathbf{K}^\top\mathbf{K}\mathbf{y} \tag{18}$$

with $\mathbf{w}_i^{(1)}$, and $\mathbf{w}_i^{(2)}$ being the i-th row of the first, and second order NN weighting matrices $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$, respectively, and the second order differencing matrix is $\Delta^{(2)}\mathbf{w}_i\mathbf{y} = \Delta\mathbf{w}_i^{(1)}\mathbf{y} - \Delta\mathbf{w}_i^{(2)}\mathbf{y}$ with $\Delta\mathbf{w}_i^{(1)}\mathbf{y} = \mathbf{w}_i^{(0)}\mathbf{y} - \mathbf{w}_i^{(1)}\mathbf{y}$, where the zero order NN weight matrix is just the original weight matrix $\mathbf{W}^{(0)} = \mathbf{W}$, and the difference is $\Delta\mathbf{w}_i^{(2)}\mathbf{y} = \mathbf{w}_i^{(1)}\mathbf{y} - \mathbf{w}_i^{(2)}\mathbf{y}$; all the neighborhood matrices $\mathbf{W}$ are partitioned row-wise: $\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 \\ \dots \\ \mathbf{w}_n \end{pmatrix}$.

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 (18).

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:

$$\mathbf{K} = \begin{pmatrix} \mathbf{w}_1^{(0)} & -2\mathbf{w}_1^{(1)} & \mathbf{w}_1^{(2)} & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & \mathbf{w}_2^{(0)} & -2\mathbf{w}_2^{(1)} & \mathbf{w}_2^{(2)} & 0 & ... & 0 & 0 & 0 \\ & ... & & ... & ... & ... & ... & ... & \\ 0 & 0 & 0 & 0 & 0 & ... & w_n^{(0)} & -2\mathbf{w}_n^{(1)} & \mathbf{w}_n^{(2)} \end{pmatrix} (\mathbf{1}_n \otimes \mathbf{I}_n) = \begin{pmatrix} \mathbf{w}_1^{(0)} - 2\mathbf{w}_1^{(1)} + \mathbf{w}_1^{(2)} \\ \mathbf{w}_2^{(0)} - 2\mathbf{w}_2^{(1)} + \mathbf{w}_2^{(2)} \\ ... \\ \mathbf{w}_n^{(0)} - 2\mathbf{w}_n^{(1)} + \mathbf{w}_n^{(2)} \end{pmatrix} : n \times n \quad (19)$$

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 $\mathbf{w}_i^{(d)}$ terms. Now we can get a HP smoother for spatial (cross-sectional) data sets in similar way as for time series using the smoothed squared loss function.

Theorem 2 (The classical spatial HP filter).
Consider the SAR model (16) and the spatial smoothness prior (18) based on distances and the the SSL (smoothed squared loss) function in (1) based on the second order differencing matrix $\mathbf{K} : n \times n$ as in (19).

The spatial HP smoother is obtained by minimizing the quadratic form in $\boldsymbol{\tau}$, where we rewrite (3) with $\mathbf{y} = [y_1, ..., y_n]^\top$, $\boldsymbol{\tau} = [\tau_1, ..., \tau_n]^\top$ and with $\mathbf{R} = \mathbf{I}_n - \rho\mathbf{W}$ as

$$\min_{\boldsymbol{\tau}} \; (\mathbf{Ry} - \boldsymbol{\tau})^\top(\mathbf{Ry} - \boldsymbol{\tau}) + \lambda\boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau} \quad (20)$$

and the solution to the optimisation problem is attained at the posterior mean

$$\boldsymbol{\tau}_{**} = [\mathbf{R}^\top\mathbf{R} + \lambda\mathbf{K}^\top\mathbf{K}]^{-1}\mathbf{Ry}. \quad (21)$$

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

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

$$(\mathbf{Ry} - \boldsymbol{\tau})^\top(\mathbf{Ry} - \boldsymbol{\tau}) + \lambda\boldsymbol{\tau}^\top\mathbf{K}^\top\mathbf{K}\boldsymbol{\tau} = (\boldsymbol{\tau} - \boldsymbol{\tau}_{**})^\top(\boldsymbol{\tau} - \boldsymbol{\tau}_{**}) + \mathbf{y}^\top\lambda\mathbf{K}^\top\mathbf{K}(\lambda\mathbf{K}^\top\mathbf{K} + \mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{Ry} \quad (22)$$

with the posterior mean $\boldsymbol{\tau}_{**} = \mathbf{A}_{**}^{-1}\mathbf{R}^\top\mathbf{y}$ and the posterior precision matrix $\mathbf{A}_{**} = [\mathbf{R}^\top\mathbf{R} + \lambda\mathbf{K}^\top\mathbf{K}]$ as in Theorem 1. If necessary, the point predictor for the spatial HP smooth is given by the posterior mean $\boldsymbol{\tau}_{**}$.

5. The extended regression and smoothing model

In this section we extend the smoothing model (1) to a general regression framework, where the additional regressors either control for other (ideosyncratic) influences or are the focus after the elimination of the HP trend:

$$\mathbf{y} = \mathbf{I}_T\boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{I}_T]. \tag{23}$$

The mean of this model is defined with $\mathbf{Z} = [\mathbf{I}_T : \mathbf{X}]$ and $\boldsymbol{\gamma}^\top = (\boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top)$ by

$$\boldsymbol{\mu} \quad = \quad \mathbf{I}_T\boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} = [\mathbf{I}_T : \mathbf{X}]\gamma = \mathbf{Z}\gamma. \tag{24}$$

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 Definition 1

$$\min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau}) \quad with \quad SSL(\boldsymbol{\tau}) = \sum_{t=1}^{T}(y_t - \mu_t)^2 + \lambda\sum_{t=1}^{T}(\Delta^2\mu_t)^2 \tag{25}$$

with $\Delta^2\mu_t = \Delta\mu_t - \Delta\mu_{t-1}$ and from (24) we get

$$\Delta\mu_t = \mu_t - \mu_{t-1} = \tau_t - \tau_{t-1} + (x_t - x_{t-1})\beta, \quad for \quad t = 1, ..., T, \tag{26}$$

5.1. The Bayesian extended HP smoothness model

In this section we discuss the extended HP (eHP) smoothing model 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 (23). 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* $\hat{\mathbf{y}}_\beta$*:*

$$\hat{\mathbf{y}}_\beta = \min_{\boldsymbol{\tau}} SSL(\boldsymbol{\tau} \mid \boldsymbol{\beta}) \quad with \quad SSL(\boldsymbol{\tau} \mid \boldsymbol{\beta}) = ESS(\boldsymbol{\tau} \mid \boldsymbol{\beta}) \ + \ \lambda * smooth(\boldsymbol{\tau}) \tag{27}$$

where $smooth(\boldsymbol{\tau})$ *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 $\hat{\mathbf{u}}_{HP} \neq \hat{\mathbf{y}}_{eHP}$ and 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 $\boldsymbol{\beta}$. With additional regressors $\mathbf{X}$ we assume for the stacked $\boldsymbol{\gamma}$ parameter we a conjugate normal-gamma model.

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

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}_T], \quad \textit{or} \quad \mathbf{y} \sim \mathcal{N}[\mathbf{Z}\boldsymbol{\gamma}, \sigma^2 \boldsymbol{\Sigma}_0], \tag{28}$$

with $\boldsymbol{\gamma} = \begin{pmatrix} \tau \\ \beta \end{pmatrix}$ *and where* $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$ *is a known covariance matrix. As prior we use the conjugate 'multivariate NG' distribution*

$$(\boldsymbol{\gamma}, \sigma^{-2}) \sim \mathcal{N}_{n+p}\Gamma[\boldsymbol{\gamma}_*, \widetilde{\mathbf{A}}_*, \sigma_*^2, n_*], \quad \textit{with} \quad \widetilde{\mathbf{A}}_* = diag(\lambda \mathbf{K}^\top \mathbf{K}, n_2 \mathbf{I}_p)^{-1} = \begin{pmatrix} (\lambda \mathbf{K}^\top \mathbf{K})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p / n_2 \end{pmatrix} \tag{29}$$

that consists of 2 independent blocks for $\boldsymbol{\beta}$ *and* $\boldsymbol{\tau}$*.*

λ is the large hypothetical sample size for the $\boldsymbol{\tau}$ parameter and n_2 is the hypothetical sample size for the $\boldsymbol{\beta} : p \times 1$ regression coefficients and for the rather non-informative prior information it could be rather small, like $n_2 = 1$. This set-up allows a Bayesian inference with conjugate normal-gamma distributions:

Theorem 3 (The conjugate extended HP smoothing model).
The conjugate Bayesian inference of the extended HP smoothing model in (28) with parameters $\theta = (\boldsymbol{\gamma}, \sigma^{-2})$ *as in Definition 3 is:*

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

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

and the likelihood of the data

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

yields the posterior distribution

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

with the parameters

$$
\begin{aligned}
\boldsymbol{\gamma}_{**} &= \widetilde{\mathbf{A}}_{**}(\widetilde{\mathbf{A}}_{*}\boldsymbol{\gamma}_{*} + \boldsymbol{\Sigma}_0^{-1}\mathbf{y}), \\
\widetilde{\mathbf{A}}_{**}^{-1} &= \widetilde{\mathbf{A}}_{*}^{-1} + \boldsymbol{\Sigma}_0^{-1}, \\
n_{**} &= n_* + n, \\
n_{**}s_{**}^2 &= n_* s_*^2 + ns^2 + \alpha \\
\alpha &= \mathbf{y}^\top \widetilde{\mathbf{A}}_{*}(\widetilde{\mathbf{A}}_{*} + \boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_0\mathbf{y}
\end{aligned} \tag{30}
$$

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 3. *Is given in Polasek (2011).*

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

For the Bayesian eHP model we specify a prior distribution for the parameters as in (23):

$$
\mathbf{K}\boldsymbol{\tau} \sim \mathcal{N}[\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_T], \quad \boldsymbol{\beta} \sim \mathcal{N}[\boldsymbol{\beta}_*, \mathbf{H}_*], \quad \sigma^{-2} \sim \Gamma[\sigma_*^2 n_*/2, n_*/2]. \tag{31}
$$

The estimation of the extended HP model (23) can be done conveniently by a MCMC procedure.

Theorem 4 (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 (23) with prior (31) 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: (33), (35) and (36).

Proof 4. *The full conditional distributions (fcd) are:*

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

$$
\begin{aligned}
p(\boldsymbol{\beta} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \mathcal{N}[\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_T] \\
&= \mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_{**}, \mathbf{H}_{**}]
\end{aligned} \tag{32}
$$

with the parameters

$$
\begin{aligned}
\mathbf{H}_{**}^{-1} &= \mathbf{H}_*^{-1} + \sigma^{-2}\mathbf{X}^\top\mathbf{X}, \\
\mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}_*^{-1}\mathbf{b}_* + \sigma^{-2}\mathbf{X}^\top(\mathbf{y} - \boldsymbol{\tau})]
\end{aligned} \tag{33}
$$

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

$$
p(\sigma^{-2} \mid \boldsymbol{\tau}, \mathbf{y}) \propto \Gamma\left[\sigma^{-2} \mid n_{**}s_{**}^2, n_{**}/2\right] \tag{34}
$$

is a gamma distribution with the parameters $n_{**} = n_* + n$

$$n_{**}s_{**}^2 = n_* s_*^2 + \sum_{i=1}^{n} (y_i - \tau_i - \mathbf{x}_i \boldsymbol{\beta})^2 \quad (35)$$

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

$$\begin{aligned} p(\boldsymbol{\tau} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\tau} \mid \mathbf{0}, \mathbf{A}_*]\, \mathcal{N}[\mathbf{y} \mid \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T] \\ &= \mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\tau}_{**}, \mathbf{A}_{**}] \end{aligned} \quad (36)$$

with the parameters $\boldsymbol{\tau}_{**} = \mathbf{A}_{**}\mathbf{y}$ *and* $\mathbf{A}_{**}^{-1} = \widetilde{\mathbf{A}}_*^{-1} + \sigma^{-2}\mathbf{X}^\top \mathbf{X}$.

6. Regional extended HP filtering of GDP and employment

In this section we show how the spatial HP model can be applied to smooth the regional GDP and employment data 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. The model we have specified is an extended HP model $\mathbf{y} = \boldsymbol{\tau} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ contains the dummy variables for the 25 EU countries to catch the fixed effects plus an extra dummy variable *Dagg* for regions that are city agglomerations. The car driving times were obtained by own calculations based on pairwise queries by internet search machines and are used for the $\mathbf{W}$ matrix and to define the smoothing metric.

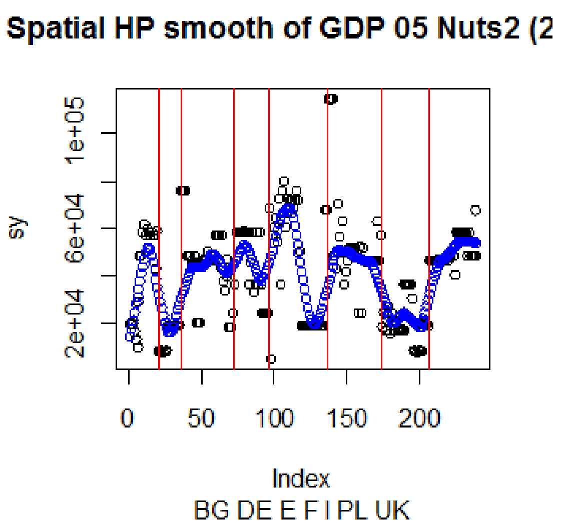
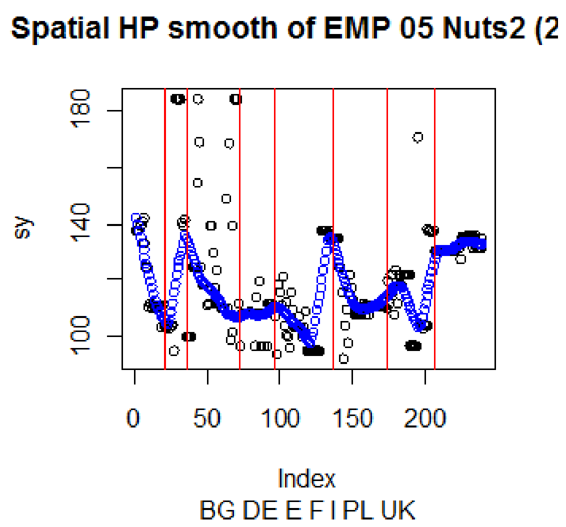


Figure 1: Spatial HP smooth of GDP 05, NUTS-2, 2005



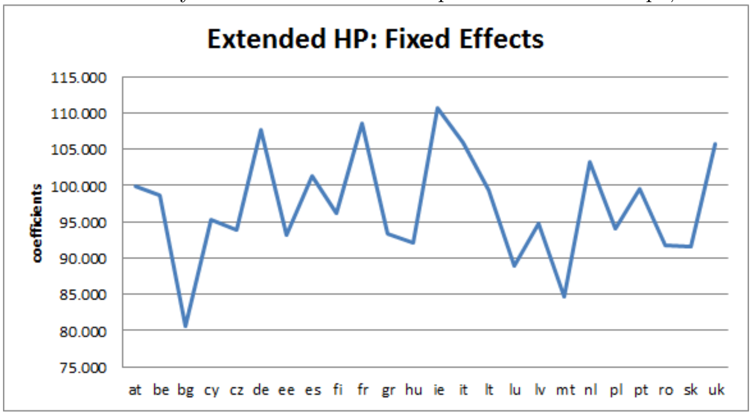
Spatial HP smooth of Employment, NUTS-2, 2005

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
```

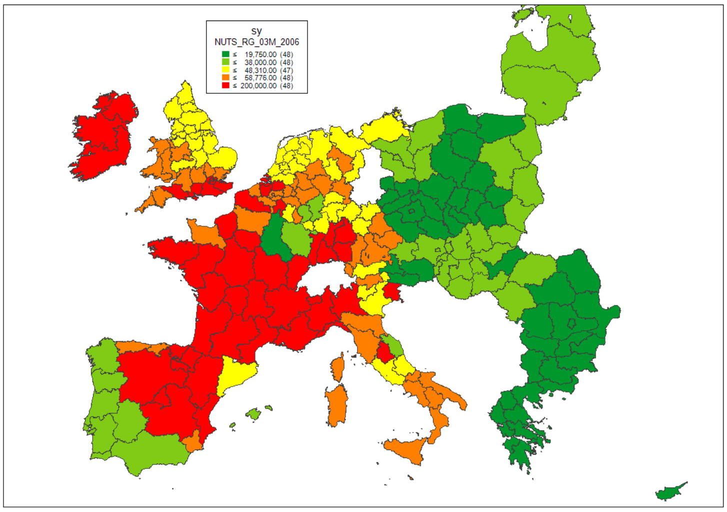
The effect of the spatial smoothing is seen in alphabetical order of the 24 countries[3] in Figure 1. The volatility of the smooth can be attributed to the heterogeneity between and within the countries. The median fixed effects coefficients of the extended spatial HP procedure were estimated with MCMC and are shown in Figure 2. These are the median effects of the 25 country dummy variables in **X**: The smallest one is Portugal and the largest one is Malta. The geographical maps for the smoothed GDP and GDPpc of NUTS-2 regions are given in the Figure 3 and in Figure 4, respectively, together with the observed raw values.

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




```
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 ***
```

[3] AT BE BG CY CZ DE EE E FI F GR HU IE I LT LU LV MT NL PL PT RO SK UK without DK SE SL



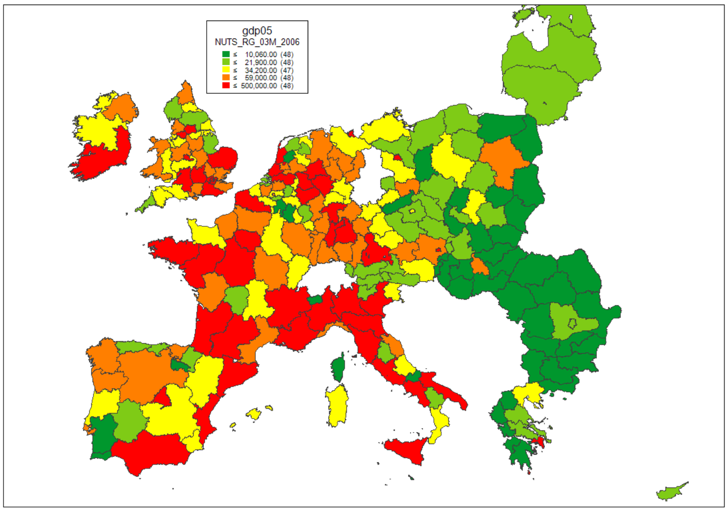
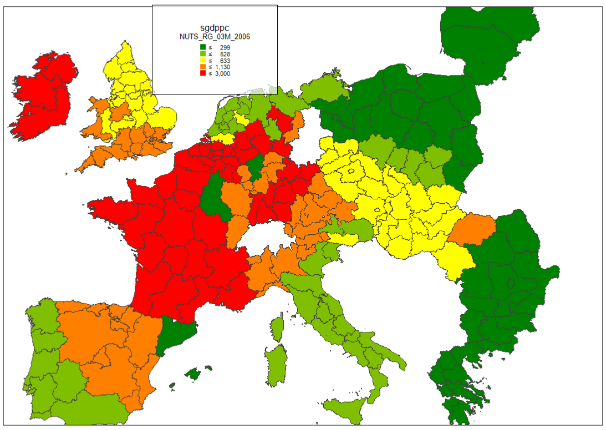


Figure 3: Spatial HP smooth of GDP NUTS-2, 2005 Map of 239 GDP NUTS-2 regions, 2005 (raw data)



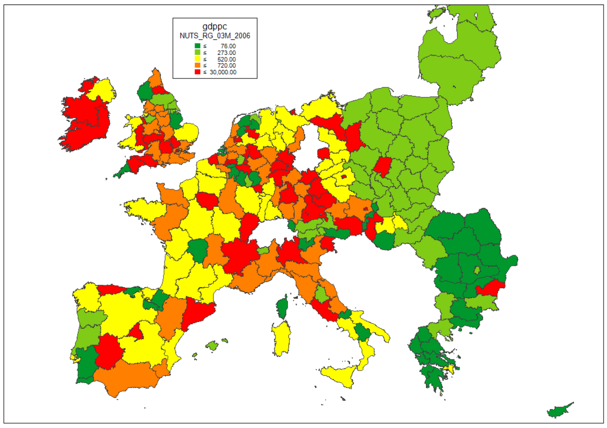


Figure 4: Spatial HP smooth of GDPpc, NUTS-2, 2005 239 GDPpc NUTS-2 regions, 2005 (raw data)

```
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
```

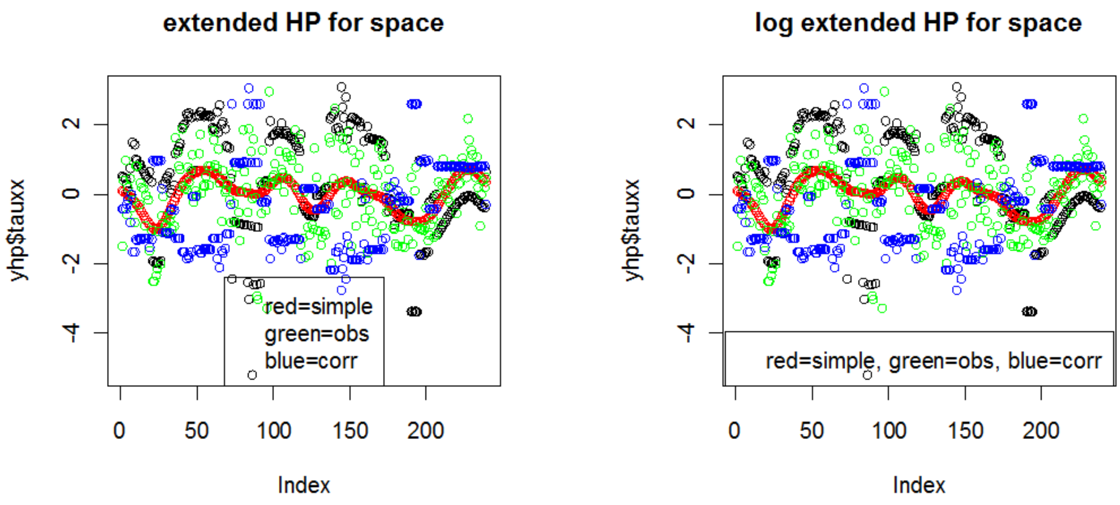


Figure 5: Spatial extended HP smooth of GDP Nuts-2, 2005   Spatial extended HP smooth of log GDP Nuts-2, 2005

6.1. Employment

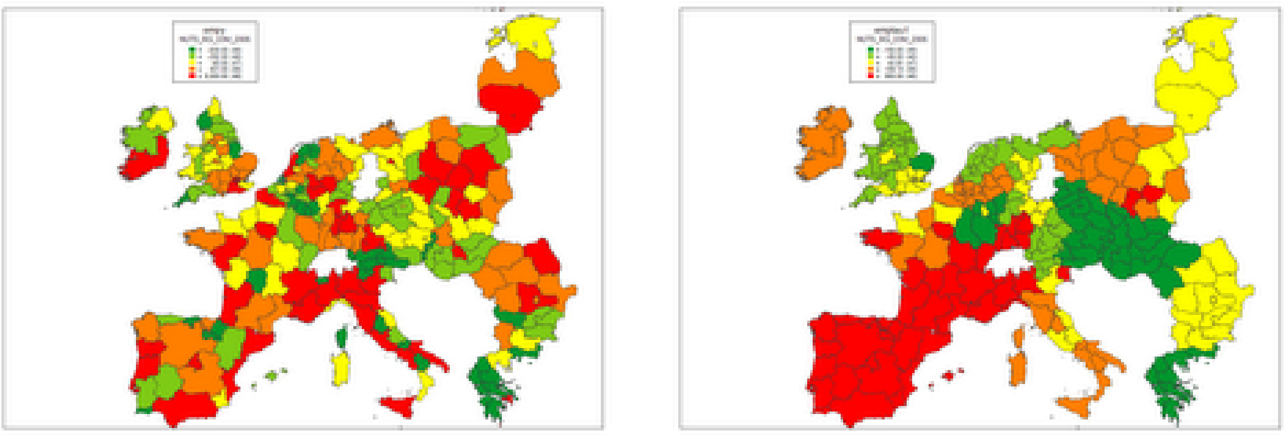
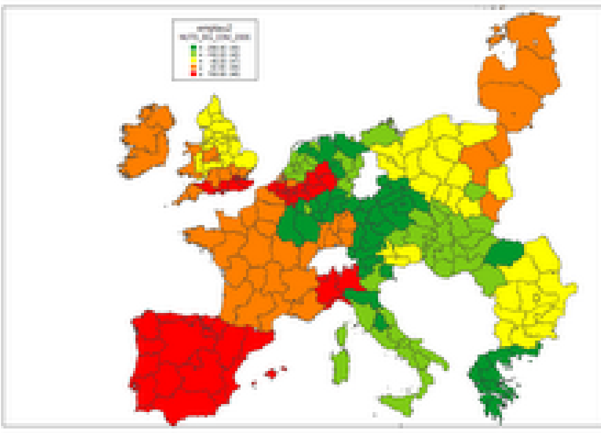


Figure 6: Employment: NUTS-2, 2005 (raw data)   MCMC Spatial HP smooth of Employment NUTS-2, 2005

Figure 6 shows the raw data together with the smooth of the employment data in 2005: the first things to note are the high employment effects in central Poland and Romania. The smooth in Figure 6 shows the smooth (posterior mean) of the spatial HP model while Figure 7 shows the smooth (posterior mean) of the spatial extended HP model. The **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: MCMC of the spatial extended HP model, smooth of Employment NUTS-2, 2005



7. Model selection and Bayes testing

This section shows how to compute the Bayes factor for the HP smoother and to select the order of smoothness prior by marginal likelihoods. The assumption to do this requires a normal prior distribution with full rank. The first order differencing matrix is denoted as $\mathbf{K}_1$ and 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 conjugate eHP model).

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

$$p(\mathbf{y} \mid eHP) = (\pi)^{-\frac{n}{2}} \frac{|\widetilde{\mathbf{A}}_{**}|^{\frac{1}{2}}}{|\widetilde{\mathbf{A}}_{*}|^{\frac{1}{2}}} \times \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_{*}}{2})} \times \frac{(n_* s_*^2/2)^{\frac{n_*}{2}}}{(n_{**} s_{**}^2/2)^{\frac{n_{**}}{2}}}, \tag{37}$$

where n_{**} and s_{**}^2 are the posterior parameter given in (30) of Theorem 3.

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

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

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

$$lml_{eHP} = -\frac{n}{2} log(\pi) + log(R_{det}) + log(R_{df}) + log(R_{ESS}). \tag{38}$$

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

$$-2lml_{HP} = nlog(\pi) + log\frac{|\widetilde{\mathbf{A}}_{**}^{-1}|}{|\widetilde{\mathbf{A}}_{*}^{-1}|} + (n_{**}-3)log(2) + 2log(\Gamma(n_{**}-1)) + n_* log(ESS_*) - n_{**} log(ESS_{**}). \tag{39}$$

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

$$R^2_{det} = \mid \widetilde{\mathbf{A}}_* \mid / \mid \widetilde{\mathbf{A}}_{**} \mid = \frac{\mid \mathbf{A}_* \mid n_2^p}{\mid \mathbf{A}_{**} \mid \mid \mathbf{G}^{-1} \mid}$$

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

Proof 5. *The determinant of the prior is* $\mid \widetilde{\mathbf{A}}_* \mid = \mid \mathbf{A}_* \mid n_2^p$ *while for the determinant of the partitioned posterior we find*[4] $\mid \widetilde{\mathbf{A}}_{**} \mid = \mid \mathbf{A}_* \mid \mid \mathbf{G}^{-1} \mid$ *with* $\mathbf{G}$ *depending on the i-th differencing matrix:* $\mathbf{G}_i^{-1} = n_2 \mathbf{I}_p + \mathbf{X}^\top \mathbf{P}_{i,\lambda} \mathbf{X}$ *with* $\mathbf{P}_{i,\lambda}$ *as before. The ratio of determinants for differencing matrices of order* i *is*

$$R^2_{det,i} = \frac{\mid \widetilde{\mathbf{A}}_{i*}^{-1} \mid}{\mid \widetilde{\mathbf{A}}_{i**}^{-1} \mid} = \frac{\mid \lambda \mathbf{K}_i^\top \mathbf{K}_i \mid}{\mid \mathbf{I}_n + \lambda \mathbf{K}_i^\top \mathbf{K}_i \mid} \frac{n_2^p}{\mid \mathbf{G}_i \mid}$$

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*

$$R_{df} = \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_*}{2})} = \frac{(n_{**} - 2)!!}{2^{(n_{**}-1)/2}} \tag{40}$$

The log df-ratio is

$$\begin{aligned} log(R_{df}) &= log \frac{(n_{**} - 2)!!}{2^{(n_{**}-1)/2}} = (n_{**} - 2) log(2) + log((n_{**} - 2)!) - \frac{(n_{**} - 1)}{2} log(2) = \\ &= \frac{(n_{**} - 3)}{2} log(2) + log(\Gamma(n_{**} - 1)), \end{aligned} \tag{41}$$

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

[4] using $\mid \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} \mid = \mid \mathbf{E} \mid \mid \mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F} \mid$

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

For the Bayes test between two eHP 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

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

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:

$$\begin{aligned} \alpha_i &= \mathbf{y}^\top n' \mathbf{K}_i^\top \mathbf{K}_i (n' \mathbf{K}_i^\top \mathbf{K}_i + \mathbf{\Sigma}_0)^{-1} \mathbf{\Sigma}_0 \mathbf{y} = \\ &= \mathbf{y}^\top (\mathbf{C}^\top \mathbf{C}/n' + \mathbf{R}^\top \mathbf{R})^{-1} \mathbf{y} \end{aligned} \quad (42)$$

with $\mathbf{C}_i = \mathbf{K}_i^{-1}$ and $\mathbf{\Sigma}_0 = (\mathbf{R}^\top \mathbf{R})^{-1}$. 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

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

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

The marginal likelihood for models $\mathcal{M}$ that are estimated by MCMC can be computed by the Newton-Raftery formula

$$\hat{m}(\mathbf{y} \mid \mathcal{M})^{-1} = \frac{1}{n_{rep}} \sum_{j=1}^{n_{rep}} \left(\sum_{i=1}^{n} ln\ l(\mathcal{D}_i \mid \mathcal{M}, \theta_j) \right)^{-1} l(\mathcal{D}_i \mid \mathcal{M}, \theta)^{-1} \quad (43)$$

where $\mathcal{D}_i = (y_i, \mathbf{x}_i)$ is the i-th data observation and with the likelihood given by the HP model (23).

8. Summary

This paper has shown that the extended HP filter can be also used to smooth spatial regional data. HP filtering 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 (of the prior information) for the smooth component τ, i.e. the parameter vector to be estimated. For the remaining coefficient we assume a small hypothetical sample size to express diffuse knowledge.

The Bayesian view of the extended 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 procedure can be applied. 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 (2011a).

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. Possibly, 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. The Bayesian formulation allows 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.

9. Appendix: The Combination of Quadratic Forms

The standard result for combining quadratic forms in normal Bayes models is:

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*

$$(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{H}(\boldsymbol{\beta}-\mathbf{b})+(\boldsymbol{\beta}-\mathbf{b}_*)^\top\mathbf{H}_*(\boldsymbol{\beta}-\mathbf{b}_*)$$
$$=(\boldsymbol{\beta}-\mathbf{b}_{**})^\top\mathbf{H}_{**}(\boldsymbol{\beta}-\mathbf{b}_{**})+(\mathbf{b}-\mathbf{b}_*)^\top\mathbf{H}_*(\mathbf{H}_*+\mathbf{H})^{-1}\mathbf{H}(\mathbf{b}-\mathbf{b}_*) \tag{44}$$

with the parameters $\mathbf{H}_{**}=\mathbf{H}_*+\mathbf{H}$ *and* $\mathbf{b}_{**}=\mathbf{H}_{**}^{-1}(\mathbf{H}_*\mathbf{b}_*+\mathbf{H}\mathbf{b})$.

10. References

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

Hodrick, R, J. and E. C. Prescott (1980), Postwar U.S. Business Cycles: an Empirical Investigation, Discussion Paper no. 451, Carnegie Mellon University. Appeared (1997).

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

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.

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