Semi-parametric Regression under Model Uncertainty: Economic Applications*

GERTRAUD MALSINER-WALLI,† PAUL HOFMARCHER,‡ and BETTINA GRÜN§

†Department of Finance, Accounting and Statistics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020, Vienna, Austria (e-mail: gertraud.malsiner-walli@wu.ac.at)
‡Salzburg Centre of European Union Studies (SCEUS), Department of Business, Economics and Social Theory, Paris Lodron University of Salzburg, Mönchsberg 2a, 5020, Salzburg, Austria (e-mail: paul.hofmarcher@sbg.ac.at)
§Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040, Linz, Austria (e-mail: bettina.gruen@jku.at)

Abstract

Economic theory does not always specify the functional relationship between dependent and explanatory variables, or even isolate a particular set of covariates. This means that model uncertainty is pervasive in empirical economics. In this paper, we indicate how Bayesian semi-parametric regression methods in combination with stochastic search variable selection can be used to address two model uncertainties simultaneously: (i) the uncertainty with respect to the variables which should be included in the model and (ii) the uncertainty with respect to the functional form of their effects. The presented approach enables the simultaneous identification of robust linear and nonlinear effects. The additional insights gained are illustrated on applications in empirical economics, namely willingness to pay for housing, and cross-country growth regression.

I. Introduction

Model uncertainty is pervasive in empirical economics. Economic theory does not always specify or is inconclusive with respect to (i) the covariates which are part of the model, and (ii) the functional form between dependent and explanatory variables. Bayesian model averaging (henceforth BMA) has become a popular tool to perform robust inference under model uncertainty (Hofmarcher et al., 2018). The model uncertainty usually accounted for in the BMA framework only relates to variable uncertainty while assuming that the relationship between dependent and explanatory variables is linear. For instance, in the field of long-term per capita GDP growth empirics a huge strand of literature applies BMA

JEL Classification numbers: C14, O10, O40.

*This research was supported by the Oesterreichische Nationalbank (Anniversary Fund, project number: 14663) and the Austrian Science Fund (FWF, P28740).

Electronic copy available at: https://ssrn.com/abstract=3615211
methods to identify robust linear determinants for growth. Variable uncertainty arises in this context due to the so-called *Open-endedness* of theory (see Brock and Durlauf, 2001) where competing growth theories do not rule out each other. BMA takes this variable uncertainty into account (see, e.g. Fernández, Ley and Steel, 2001; Sala-i-Martin, Doppelhofer and Miller, 2004; Amini and Parmeter, 2012; Ley and Steel, 2012). BMA techniques are also employed in Mitchell, Pain and Riley (2011) to identify economic and policy drivers of international migration to the UK, in Eicher, Henn and Papageorgiou (2012) to assess effects of preferential trade agreements on trade flows, and in Tobias and Li (2004) to determine returns to education.

Although the insights from these BMA exercises are valuable, they rely on the assumption that the functional relationship between the potential covariates and the dependent variable is known. In standard BMA all covariates included in the model are assumed to have a linear effect on the outcome. A nonlinear functional relationship is modelled by manually creating and adding new covariates, e.g. by including polynomial terms (Henderson and Parmeter, 2016) or defining a piecewise linear function (Tobias and Li, 2004). Special care is needed if these additional terms are to be included in the model in a hierarchical or grouped way. For instance, if interaction effects are only to be included jointly with their main effects, the constraint has to be imposed that main effects need to be included first before interaction effects can be included (Crespo Cuaresma, 2010; Montgomery and Nychka, 2010). The inclusion of grouped terms representing a nonlinear effect requires that all terms needed to model nonlinearity are either jointly added or dropped from the model. Due to this manual tuning required for the inclusion of any nonlinear effect, BMA applications in general only include a small subset of potential covariates with nonlinear effects.

However, functional misspecification occurs if covariates are only included with linear effects while in fact their effects are actually nonlinear. In this case, several issues may arise which preclude the correct robust identification of effects: (i) irrelevant covariates may be included into the model to compensate for the missing nonlinear effects and are overestimated in terms of importance, (ii) important nonlinear covariates may not be identified due to the functional misspecification, and (iii) effects of covariates, which vary over the covariate range, are forced to be constant over the entire covariate range. Inference based on an improperly specified model thus may lead to incorrect conclusions and policy prescriptions.

Therefore, efforts have been made in empirical economics to develop and apply modelling techniques which uncover also nonlinear effects without the need to know and a priori explicitly define these nonlinear effects. Non-parametric as well as semi-parametric methods have been suggested. Non-parametric methods used in economic regression applications are mainly based on kernel estimation (for an overview see Henderson and Parmeter, 2015). For instance, local-linear least square regressions are used in Henderson, Papageorgiou and Parmeter (2011) to investigate nonlinearities of different economic growth determinants and in Delgado, Henderson and Parmiter (2014) to study the effect of education on economic growth. Semi-parametric methods using model matrix expansions to obtain more flexible regression functions are discussed in Smith and Kohn (1996) to model housing values, and in Koop and Tobias (2006) to study returns to education.

Compared to BMA, where dozens of potential determinants are simultaneously included in the analysis to identify the relevant ones, the non- and semi-parametric methods...
considered so far were only used with a limited subset of covariates. Henderson et al. (2011) consider at most 11, Smith and Kohn (1996) five and Henderson and Parmeter (2016) two nonlinear regressors. Thus these approaches account for functional uncertainty, but only after having restricted the potential set of covariates and relying on prior knowledge to reduce variable inclusion uncertainty. The results obtained using these approaches therefore do not account for the complete model uncertainty. This shortcoming is also pointed out by Henderson et al. (2011), who state that future research should focus on merging the use of nonlinear regression methods with BMA techniques. The approach presented in this paper addresses this issue. In particular, we propose to use recently developed Bayesian variable selection methods which incorporate nonlinear effect estimation into the BMA framework using a semi-parametric approach (Scheipl, Fahrmeir and Kneib, 2012). In the following, we refer to this approach as Spike-Slab-GAM (henceforth, SSG).

The SSG approach is appealing for several reasons. In contrast to standard BMA, SSG simultaneously accounts for two sources of model uncertainty, the variable inclusion uncertainty and the functional form uncertainty of each variable. SSG estimates the functional relationships between the covariates and the dependent variable in a very flexible way through smooth functions, where their specific form does not need to be a priori specified. For each covariate, two terms are included, a linear and a nonlinear term. Each of these two terms can be selected independently and the inclusion probability of both terms can be evaluated separately. The results indicate for each covariate (i) if it is a robust linear determinant of the outcome and (ii) if it has robust nonlinear effect on the outcome. 

The smooth functions are modelled using penalized spline regression (Eilers and Marx, 1996). This alleviates the problem of specifying an appropriate spline basis by selecting a suitable order, number and location of the basis functions, and gives good results as long as a flexible enough spline basis is specified. A specifically designed prior construction on the regression coefficients regularizes the estimation of the high-dimensional parameter vector and allows simultaneous consideration of a large number of nonlinear effects. For example, when analysing the data set by Sala-i-Martin et al. (2004) in section V more than 50 nonlinear effects of covariates are simultaneously considered.

The results of SSG can be analysed and interpreted in a similar way to those obtained from standard BMA. The same key characteristics, such as posterior inclusion probabilities, are available for each linear and nonlinear term to assess the effect of a covariate. Thus, SSG identifies the robust linear and nonlinear effect of a covariate and indicates whether a covariate is a relevant linear and/or nonlinear driver of the outcome or not relevant at all.

The SSG model specification can also be modified in the same ways as suggested previously for BMA. For example, as a response to Ciccone and Jarocinski (2010) who criticize the instability of BMA results, Rockey and Temple (2016) suggest to pursue a less ‘agnostic’ approach and to restrict the model space by forcing some covariates to be included. This idea can also be used with SSG to selectively include covariates with specific functional relationships. In a similar vein, also interaction terms as considered by Montgomery and Nyhan (2010), can be added or the inclusion of bivariate nonlinear terms considered.

Spike-Slab-GAM differs from standard BMA employed in empirical economics in two ways: (i) nonlinear effects are specified with an a priori unknown functional form, and (ii) a slightly different prior on the regression coefficients is imposed to perform variable
selection. The modification of the prior structure is necessary to include or exclude all coefficients belonging to the same nonlinear effect jointly and to simultaneously estimate inclusion probabilities and the smoothness penalty in a computationally feasible way.

Our work is organized as follows: In section II, we present the details of model specification for SSG. We indicate how SSG identifies robust determinants under variable uncertainty while supporting arbitrary functional nonlinear relationships between the covariates and the outcome. More details on the methodology are provided in online Appendix B where a general statistical framework is presented and discussed which contains the BMA and SSG approaches as special cases. The importance of accounting for nonlinear effects is illustrated in a simulation study in section III, which provides a comparison between BMA using a standard implementation (Zeugner and Feldkircher, 2015) and SSG. We illustrate that SSG is superior to BMA if nonlinearities are present in the data. The simulation study also indicates that the cost associated with using SSG compared to BMA is negligible, even if the true functional relationship is linear. SSG is applied to economic applications consisting of willingness to pay for housing (Harrison Jr. and Rubinfeld, 1978) in section IV, and to cross-country growth regression (Fernández et al., 2001; Sala-i-Martin et al., 2004) in section V. In these applications, well-known data sets are reanalysed using SSG, and results are compared to BMA and a variant of SSG where only linear effects are included. The effect estimates of the covariates exhibiting nonlinear effects are visualized. The insights gained by including the nonlinear effects are pointed out. Finally, section VI concludes.

II. Methodology

This section describes the model specification of SSG as proposed by Scheipl et al. (2012) and discusses estimation, inference and postprocessing tools. First, linear additive models based on penalized splines are introduced. This flexible model class is able to fit any smooth function to represent the functional relationship between a covariate and the outcome variable. Then, the linear additive model is combined with stochastic search variable selection. This combination accounts for variable inclusion and functional relationship uncertainty. Finally, suitable posterior inference tools to assess a covariate’s importance and its robust effect are discussed.

Linear additive models with penalized splines

We assume the following regression setting of \( n \) observations with a dependent variable \( y \in \mathbb{R}^n \) and a set of explanatory variables \( x_k \in \mathbb{R}^n, k = 1, \ldots, K \), where the dependent variable is assumed to be given by

\[
y = \alpha \mathbf{1} + \sum_{k=1}^{K} f_k(x_k) + \epsilon,
\]

with \( \mathbf{1} \) a vector of ones of length \( n \), \( \alpha \) the intercept, the function \( f_k(\cdot) \) capturing the influence of each covariate \( x_k \) on the dependent variable and \( \epsilon \) a length \( n \) vector of i.i.d. normally distributed noise with mean 0 and variance \( \sigma^2 \). This kind of regression model is also referred to as a ‘linear additive model’ (Wood, 2006).
The unknown functions $f_k(\cdot)$ are assumed to have mean zero (for identifiability of the intercept) and to be smooth, but are otherwise not further specified or restricted. Different modelling approaches have been proposed to estimate $f_k(\cdot)$ including non-parametric approaches (Li and Racine, 2007) and semi-parametric methods based on splines, see, e.g. Smith and Kohn (1996) and Wood (2006).

Splines are smooth functions which approximate any kind of functional shape in a flexible way. They are used in a regression framework in case the functional form of the effects is unknown. Splines are defined by piecewise polynomials which are generated by basis functions. The spline basis expansion leads to an extended model matrix for the regression model. Using such a spline basis expansion in the regression implies that the smooth function is approximated by

$$f_k(x_k) \approx S_k \hat{\delta}_k,$$

with $S_k \in \mathbb{R}^{n \times (L_k + 1)}$ the spline basis expansion of $x_k$ with dimension $L_k + 1$ and $\hat{\delta}_k$ the regression coefficients.

The selection of the order of the polynomials, the number of knots acting as anchor points for the polynomial pieces and the location of the knots are crucial choices for the spline basis. All these parameters determine how flexible the spline function is. If the polynomial order or the number of knots is too low or the locations are not well selected the functional relationship between covariate and dependent variable cannot be estimated well, but a bias is introduced. On the other hand, if the polynomial order and the number of locations are too high, too wiggly and overfitting curves may be obtained.

For penalized splines these specifications are less crucial. Penalized splines are used in combination with a generous spline basis expansion which includes more flexibility than actually needed. The appropriate smoothness of the estimated function is then determined by imposing a penalty on the regression coefficients of the spline basis expansion. Thus, the problem of selecting the polynomial order and the number and location of the knots is reduced to the problem of selecting a suitable penalty value. This considerably facilitates the model specification because default settings can be employed for the basis expansion (Wood, 2006).

Eilers and Marx (1996) propose to use penalized cubic $B$-splines as basis functions. They exploit that $B$-splines only act locally, i.e. they are non-zero only on a sub-interval of the interval where the function is defined, and impose a penalty on the second differences of the coefficients of adjacent $B$-splines to induce smoothness of the estimated curve. Using the second differences as a penalty implies that constant and linear functions do not incur a penalty, but other local changes in functional shape are penalized. The regression coefficients $\alpha$ and $\hat{\delta}_k$, $k = 1, \ldots, K$, are then determined by minimizing the ordinary least squares (OLS) criterion plus the penalty term, i.e.

$$(y - \alpha t - \sum_{k=1}^{K} S_k \hat{\delta}_k)^\top (y - \alpha t - \sum_{k=1}^{K} S_k \hat{\delta}_k) + \sum_{k=1}^{K} \lambda_k (\Delta^2 \hat{\delta}_k)^\top (\Delta^2 \hat{\delta}_k),$$

where $\Delta^2 \hat{\delta}_k$ denotes the second differences of vector $\hat{\delta}_k$. Suitable values for the penalty parameter $\lambda_k$, $k = 1, \ldots, K$ need to be determined such that a smooth function is obtained without overfitting the data. For $\lambda_k = 0$ the fitted model corresponds to the OLS fit given
the spline basis expanded model matrix of $x_k$, while for $\lambda_k$ tending to $\infty$ the OLS fit for the model matrix including only the intercept and the linear term $x_k$ is obtained, i.e. the fitted smooth function is shrunken towards the best-fitting linear function.

The penalized estimation (2) corresponds to a Bayesian regression analysis where a normal prior on the regression coefficients is imposed (Lang and Brezger, 2004; Fahrmeir, Kneib and Konrath, 2010). The penalty parameters $\lambda_k$, $k = 1, \ldots, K$, are related to the precision parameters in these priors and a hierarchical prior enables a data-driven estimation of the smoothness penalty.

Since second-order differences lead to shrinkage towards a linear function, the constant and linear terms are not penalized in this setting. Therefore, in order to ensure a proper prior structure Scheipl et al. (2012) propose to split the smooth function given in equation (1) into two components: (i) the linear term and (ii) the nonlinear term corresponding to the part of the smooth function which is orthogonal to intercept and linear term, i.e. with mean and slope zero. The model matrix $S_k$ is thus split into two model matrices. One matrix consists of the linear term $x_k$, the other matrix $Z_k$ is the orthogonal complement to the intercept and linear term, capturing the deviation of $f_k(\cdot)$ from the linear function:

$$S_k \tilde{\delta}_k = x_k \beta_k + Z_k \delta_k,$$

where $Z_k \in \mathbb{R}^{n \times L_k}$.

Using this representation, the final regression model is given by

$$y = \alpha + \sum_{k=1}^{K} (x_k \beta_k + Z_k \delta_k) + \epsilon,$$

where $\delta_k$ needs to be suitably penalized to obtain a function with appropriate smoothness.

**Stochastic search variable selection**

The aim of SSG is to infer if $x_k$ has a linear, a nonlinear, or a smooth effect (which results from a combination of linear and nonlinear effect) on the outcome. This can be determined by estimating whether the coefficients $\beta_k$, $\delta_k$, or both, are different from zero. In a Bayesian variable selection setting, priors for $\beta_k$ and $\delta_k$ are specified to enable this identification. Variable selection priors considered in the Bayesian literature so far can be roughly classified into two groups, shrinkage priors and spike-and-slab priors.

Shrinkage priors are prior distributions with a lot of mass at and/or close to zero. This reflects the belief that the coefficients may eventually be zero and the variables omitted from the model. Examples for shrinkage priors are the Laplace prior, which corresponds to lasso estimation in a frequentist setting (Tibshirani, 1996; Park and Casella, 2008), or more recently, the Dirichlet–Laplace prior (Bhattacharya et al., 2015).

In contrast, spike-and-slab priors (Mitchell and Beauchamp, 1988) impose a two-component mixture distribution on the coefficients $\beta_k$ and $\delta_k$ for $k = 1, \ldots, K$. Both mixture components have mean zero, however, the variances differ. One component has all its mass close to zero or is even a degenerate point mass distribution at zero, called the ‘spike’, and the other component follows a flat distribution which would be used as prior if it were clear that this variable should be included into the model (the ‘slab’). For an overview on
Regression under model uncertainty

spike-and-slab priors see, for example, O’Hara and Sillanpää (2009) and Malsiner-Walli and Wagner (2011).

Both approaches, the standard BMA as well as the proposed SSG, implement special cases of a spike-and-slab prior. Compared to shrinkage priors, spike-and-slab priors have the advantage that assessment of effect importance is straightforward and can be based on inference tools developed for BMA. For instance, the importance of a regressor is determined by evaluating the posterior weight of slab assignment for this regressor, the so-called ‘posterior inclusion probability’.

In standard BMA the spike-and-slab prior has a spike distribution which consists of a degenerate distribution with point mass at zero and a slab distribution based on the normal g-prior (see, e.g. Zeugner and Feldkircher, 2015). In SSG, the spike-and-slab prior consists of a ‘parameter-expanded normal-mixture of inverse Gammas (peNMIG) prior’. This prior builds on the normal-mixture of inverse Gammas (NMIG) prior proposed by Ishwaran and Rao (2005). The prior is denoted by the following:

\[
\beta_k, \delta_k \sim \text{peNMIG}(\nu_0, w, a_\tau, b_\tau), \quad k = 1, \ldots, K,
\]

where \(a_\tau\) and \(b_\tau\) are the parameters of the inverse Gamma prior for the slab variance, \(\nu_0\) is the ratio between spike and slab variance and \((w, 1 - w)\) are the weights of the two-component mixture. Additionally, the weight \(w\) is assumed to follow a Beta distribution:

\[
w \sim \text{Beta}(a, b).
\]  

The peNMIG prior enables the joint inclusion and exclusion of coefficients of the linear terms \(\beta_k\) and the nonlinear terms \(\delta_k\) together with the determination of a suitable penalty, i.e. smoothness, for the nonlinear function.\(^1\)

In addition to the priors on the regression coefficients which enable stochastic search variable selection, priors for the intercept \(\alpha\) and the error variance \(\sigma^2\) are also needed for Bayesian inference. A standard specification is used assuming

\[
\alpha, \sigma^2 \propto \text{Gamma}^{-1}(a_\sigma, b_\sigma).
\]

Markov chain Monte Carlo (MCMC) methods are employed to perform Bayesian inference for SSG. Analogous to standard BMA, data augmentation is performed and a term inclusion indicator vector \(\gamma\) is introduced which is added as a latent variable to the sampling scheme. The binary indicator vector \(\gamma\) takes \(2^K\) different values in \(\{0, 1\}^{2K}\) and indicates for each term whether it is assigned to the slab or the spike. In standard BMA, a collapsed Gibbs sampling scheme is used where only \(\gamma\) is sampled, whereas all other parameters are integrated out. By contrast, a block-wise Gibbs sampler is used for estimating SSG where the regression coefficients are not integrated out but are also sampled in addition to the term inclusion vector \(\gamma\). A detailed description of the MCMC methods for estimating SSG are outlined in Scheipl (2011) and Scheipl et al. (2012).

\(^1\) For more details and properties of the peNMIG prior as well as the hierarchical representation including an indicator variable \(\gamma\) indicating assignment to the slab see online Appendix B. In the online appendix also a detailed comparison between BMA and SSG is given.
Posterior inference

The term inclusion indicator vector $\gamma$ indicates whether a coefficient is assigned to the slab or to the spike component. Assignment to the spike is interpreted as having the coefficient excluded from the model, while assignment to the slab implies that the same prior is imposed on the regression coefficient as if the covariate had been a priori included. Therefore, the indicator variable $\gamma$ is used to determine the importance of a linear or nonlinear term through the posterior inclusion probability (PIP), i.e., the posterior expectation of the term being assigned to the slab. Each $\gamma$ can be interpreted as a different model $M$. The posterior inclusion probability for term $k$, where $k = 1, \ldots, 2K$ comprising linear and nonlinear terms, is then given by

$$\text{PIP}_k = \sum_{j=1}^{2^K} I(\gamma_k = 1|M_j)P(M_j|y). \quad (5)$$

and is estimated by the proportion of MCMC samples for which $\gamma_k = 1$.

The posterior mean (PM) for the linear term of variable $k$ is determined by

$$\text{PM}_k = \sum_{j=1}^{2^K} P(\beta_k|M_j)P(M_j|y),$$

and the conditional positive sign certainty (PSC) for the linear term of variable $k$ by

$$\text{PSC}_k = \frac{\sum_{j=1}^{2^K} P(\beta_k > 0|M_j)P(M_j|y)}{\sum_{j=1}^{2^K} I(\gamma_k = 1|M_j)P(M_j|y)}.$$

While PIP is a measure of variable importance, PM indicates the robust estimate of the effect of the regressor across all sampled models, and PSC is a measure for the posterior confidence in the sign of this robust effect estimate being positive (Sala-i-Martin et al., 2004).

For hierarchical Bayesian models, the deviance information criterion (DIC) has been suggested as a general criterion to assess and compare the goodness-of-fit of different models (Spiegelhalter et al., 2002). The DIC can be directly calculated from the MCMC draws and is given by the posterior mean deviance plus the effective number of parameters. The effective number of parameters is estimated by the difference between the posterior mean deviance and the deviance at the posterior mean. Smaller values of the DIC indicate a better model fit.\(^2\)

III. Simulation study

In the simulation study, it is investigated whether (i) BMA and SSG perform similarly well if the assumption that the covariates affect the dependent variable linearly holds and (ii)

\(^2\)The derivation of the DIC for BMA where due to collapsed Gibbs sampling no MCMC samples are available for all parameters is given in online Appendix B.5.
SSG provides suitable results in case this assumption is violated and a nonlinear effect is present whereas BMA fails to identify the relevant covariates.

Design and specifications

Two different simulation settings are considered depending on whether only linear effects (‘linear setting’) or linear effects and an additional nonlinear effect determine the outcome (‘nonlinear setting’). For each setting 100 data sets are generated with each data set consisting of 50 observations and 35 covariates. In both settings five of the covariates have a linear effect. In the nonlinear setting an additional variable with a nonlinear effect is included. In particular, for the linear setting the dependent variable is given by

$$y_i = \sum_{k=1}^{5} x_{ik} + \epsilon_i,$$

while in the setting containing a nonlinear effect the dependent variable is given by

$$y_i = \sum_{k=1}^{5} x_{ik} - 1.5x_{i6}^2 + \epsilon_i.$$

The remaining covariates have no effect on the dependent variable. All covariates are drawn from a multivariate normal distribution with mean zero and variance one and with either (i) zero correlations or (ii) correlations equal to 0.5. The errors $\epsilon_i$ are drawn from $N(0, 1)$.

The simulation study compares the results for BMA and SSG where for SSG two model specifications are considered: (i) in “SSG-linear” only linear terms and (ii) in “SSG-smooth” linear and nonlinear terms are included. For BMA 1,000,000 iterations are performed after a burn-in of 20,000 iterations; for SSG 35,000 iterations after a burn-in of 5,000 iterations are made. Model fitting is performed using standardized covariates and dependent variable.\(^3\)

For each data set the results are evaluated by determining (i) the mean posterior model size, (ii) the number of true positives, i.e., the number of covariates among all selected covariates which truly have an effect on the outcome, (iii) the mean absolute deviation (MAD) between the true inclusion indicator and the estimated posterior inclusion probability, (iv) the mean squared error (MSE) between the true and predicted outcome, and (v) the DIC values. Aggregate results over the 100 data sets are reported by determining the median and the robust standard deviation.

The mean posterior model size is estimated by the average number of covariates which are assigned to the slab distribution. The number of true positives TP\(_j\) for model $M_j$ corresponds to the number of covariates which are contained in the true model as well as in $M_j$ (as induced by the assignments to the slab). More formally, TP\(_j\) is determined by

$$TP_j = \sum_{i=1}^{V} \left( 1 - \prod_{k \in K_i} (1 - I(y_k = 1|M_j)) \right),$$

\(^3\) More details on the software used and hyper-parameter values specified are given in online Appendix B.3.
where \( V \) is the number of variables with non-zero effects (\( V = 5 \) in the linear setting and \( V = 6 \) in the nonlinear setting) and \( K_v \) is the set of terms associated with variable \( v \). For BMA and SSG-linear \( K_v \) contains a single term, while for SSG-smooth \( K_v \) contains two terms, the linear and the nonlinear term associated with variable \( v \). To determine the overall TP criterion for one data set the median number of true positives is estimated based on all models, weighted with their posterior evidence.

The MAD criterion for variable inclusion is estimated by

\[
\text{MAD}_{\text{PIP}} = \frac{1}{K} \sum_{k=1}^{K} \| \hat{\text{PIP}}_k - \text{PIP}_{\text{true}} \|,
\]

with \( \hat{\text{PIP}}_k \) the estimated PIP of variable \( k \) and \( K = 35 \) in this simulation study. In the nonlinear setting, \( \hat{\text{PIP}}_k \) corresponds to the estimated posterior probability of including either a linear and/or a nonlinear term.

The MSE criterion to assess predictive performance is estimated by

\[
\text{MSE}_y = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2,
\]

with \( \hat{y}_i \) the predicted outcome and \( y_i \) a newly sampled outcome value given the same covariates \( x_i \) in order to avoid an in-sample bias. Predictions are obtained based on posterior mean estimates of the regression coefficients.

**Results**

Simulation outcomes are summarized in Table 1. The upper part of the table provides the results for the setting where only linear effects affect the outcome. In this case all three approaches are able to identify the correct model. The estimated models contain all relevant variables regardless of if the covariates are independent or correlated (median TP = 5) and only a small number of unrelated covariates are additionally included, as indicated by the median model sizes which are between 6.4 and 8.2. According to the MAD criterion the expected probability of wrongly excluding a relevant variable or wrongly including an irrelevant one is small for all three approaches. With respect to MSE BMA and SSG-linear seem to perform slightly better than SSG-smooth. This may be due to the additional complexity involved when fitting SSG-smooth. The DIC values suggest that all three approaches result in a similar model fit.

The lower part of Table 1 provides the results for the nonlinear setting. In this case BMA as well as SSG-linear misspecify the model, while SSG-smooth is able to identify the nonlinear effect and encompasses the correct model specification. Clearly BMA and SSG-linear perform worse than SSG-smooth regardless of the correlation structure between the covariates. In the case of uncorrelated covariates the estimated models are too small (with a median model size of 2.8 and 3.4) and hardly contain any relevant covariates (median TP = 1 and 2). This issue is less pronounced in the case of correlated covariates. However, still not all relevant covariates are included, but some irrelevant ones are contained. In the case of correlated covariates also a high variability in the mean posterior model size is
TABLE 1
Simulation results

| Setting                  | Estimation method | Model size | TP | MAD<sub>PIP</sub> | MSE<sub>y</sub> | DIC        |
|--------------------------|-------------------|------------|----|-------------------|----------------|------------|
| Linear, uncorrelated     | BMA               | 7.8 (0.9)  | 5  (0)| 0.08 (0.02)       | 0.21 (0.04)   | 71.0 (11.6)|
|                          | SSG-linear        | 6.4 (0.5)  | 5  (0)| 0.04 (0.01)       | 0.21 (0.04)   | 73.6 (10.2)|
|                          | SSG-smooth        | 8.1 (0.7)  | 5  (0)| 0.08 (0.02)       | 0.23 (0.05)   | 74.8 (11.1)|
| Linear, correlated       | BMA               | 7.6 (1.1)  | 5  (0)| 0.08 (0.03)       | 0.23 (0.05)   | 71.1 (11.3)|
|                          | SSG-linear        | 6.4 (0.4)  | 5  (0)| 0.04 (0.01)       | 0.23 (0.05)   | 74.5 (9.9) |
|                          | SSG-smooth        | 8.2 (0.7)  | 5  (0)| 0.08 (0.02)       | 0.25 (0.05)   | 74.9 (10.8)|
| Nonlinear, uncorrelated  | BMA               | 2.8 (0.9)  | 1  (1)| 0.16 (0.01)       | 0.63 (0.09)   | 140.7 (4.1)|
|                          | SSG-linear        | 3.4 (0.6)  | 2  (0)| 0.15 (0.01)       | 0.56 (0.05)   | 134.1 (4.6)|
|                          | SSG-smooth        | 9.2 (0.5)  | 6  (0)| 0.06 (0.01)       | 0.12 (0.02)   | 50.9 (11.3)|
| Nonlinear, correlated    | BMA               | 6.2 (2.6)  | 4  (1)| 0.14 (0.01)       | 0.37 (0.11)   | 119.2 (15.1)|
|                          | SSG-linear        | 5.3 (1.2)  | 4  (1)| 0.11 (0.02)       | 0.39 (0.09)   | 118.9 (10.2)|
|                          | SSG-smooth        | 8.6 (0.7)  | 6  (0)| 0.06 (0.01)       | 0.15 (0.03)   | 56.3 (10.5)|

Notes: Median values with robust standard deviations in parentheses over 100 data sets are given for each criterion. The correct number of true positives (TP) is five in the linear and six in the nonlinear setting.

Figure 1. Simulation results, nonlinear data set, uncorrelated covariates: Estimated effect of variable x<sub>6</sub> using SSG-linear (on the left) or SSG-smooth (on the right).

observed, as indicated by the robust standard deviations which are 2.6 for BMA and 1.2 for SSG-linear. This inferior performance is also reflected by the higher values for MAD<sub>PIP</sub> and MSE<sub>y</sub> compared to SSG-smooth. SSG-smooth identifies all relevant covariates and also provides good PIP estimates and predictive performance. The DIC values indicate that SSG-smooth provides a considerably better model fit than BMA and SSG-linear.

For one nonlinear uncorrelated data set, Figure 1 shows the estimated effect of covariate x<sub>6</sub>, if SSG-linear (on the left) and SSG-smooth (on the right) are used for model fitting. The posterior medians of the smooth effects are indicated by the black lines as well as the 80% confidence bands indicated by the grey shaded areas. Clearly for SSG-linear, where
the effect is restricted to be linear, no impact of \( x_6 \) on the dependent variable is detected, while the nonlinear effect is identified when using SSG-smooth.

In the simulation study BMA and SSG-linear give similar results despite implementing a different variant of a spike-and-slab prior and imposing different priors on the regression coefficients. Furthermore, SSG-smooth provides a suitable approach to identify robust determinants and achieve good predictive performance when a nonlinear relationship between the covariates and the dependent variable cannot be safely a priori excluded. Also, the cost of switching to this approach is small if the linear model constitutes the correct specification. The use of SSG-smooth safeguards against obtaining completely unreliable results if the assumption of a linear relationship is violated.

IV. Application to the Boston housing data set

Smith and Kohn (1996) used the Boston housing data set to indicate the advantages of selectively including nonlinear terms when predicting median home values from geographical and infrastructure characteristics of the property while accounting for model uncertainty. Thus, the aim in the following is to replicate the nonlinear effects reported by Smith and Kohn (1996) in an analysis where all variables are considered to have effects that are potentially nonlinear.

Results are presented for BMA, SSG-linear and SSG-smooth. This enables a comparison between BMA and SSG-linear to assess the impact of the different prior constructions. The results for SSG-smooth then provide insights into the differences obtained due to the additional inclusion of nonlinear effects. The results also indicate the usefulness of SSG compared to BMA in detecting nonlinear relationships.

Design and specifications

Before analysis the data set was standardized. This ensures that a similar amount of shrinkage in terms of standard deviations of the covariates is imposed on each of the covariates. BMA was run with 2,000,000 burn-in iterations and 2,000,000 recorded iterations, while SSG was run with 10 parallel chains using 2,000,000 burn-in iterations and 200,000 iterations recorded with a thinning of 5 for each chain. The same estimation methods and hyper-parameter values are used as in the simulation study in section III.4

The results for BMA, SSG-linear and SSG-smooth are assessed by a detailed comparison of the estimated PIPs, PMs and PSCs for the linear terms and the estimated PIPs for the nonlinear terms. In addition the posterior model size distributions of SSG-linear and SSG-smooth are compared where the model size is determined by the number of included linear and nonlinear terms. The posterior model size distribution is further inspected for SSG-smooth by splitting the number of included terms into the number of linear terms, the number of nonlinear terms without the linear term of this variable included and the number of nonlinear terms with the linear term of this variable included. For the variables with the highest posterior evidence for a nonlinear effect the smooth effects, i.e. the combined linear and nonlinear effects of a covariate, are visualized by their 80% confidence bands

4 More details on the specified prior parameters are given in online Appendix B.3.

© 2019 The Authors. Oxford Bulletin of Economics and Statistics published by Oxford University and John Wiley & Sons Ltd.

Electronic copy available at: https://ssrn.com/abstract=3615211
using the corresponding 10% and 90% quantiles of the posterior distributions as well as by the median of the posterior distributions.

Results

Harrison Jr. and Rubinfeld (1978) investigated the willingness to pay for clean air using data for the Boston metropolitan area. This Boston housing data set was also used by Smith and Kohn (1996) to present their approach where the functional relationship between a small subset of selected covariates and the dependent variable was specified as potentially nonlinear. The data set consists of 506 observations on 13 covariates to predict the median price of owner-occupied homes in USD 1000’s based on data from the 1970 census. In contrast to Smith and Kohn (1996), where only five covariates are modelled nonlinearly, in SSG-smooth all covariates (except for the Charles river dummy) are included with linear and nonlinear effects.

Table 2 reports the results for BMA, SSG-linear and SSG-smooth. For the Charles River dummy (chas) no nonlinear effect is estimated. This is indicated in the column PIP for the nonlinear effects by a dash (–). The results for BMA and SSG-linear are very similar in terms of PIP, PM and PSC. Both identify nearly all covariates as being robust determinants for the willingness to pay for housing. Only proportion of non-retail business acres per town (indus) and the proportion of owner-occupied units built prior to 1940 (age) have comparably small PIPs. If SSG-smooth is used and nonlinear terms are also included in the model, the set of variables identified as robust determinants is considerably reduced. In particular the variables proportion of residential land zoned for lots over 25,000 sq.ft. (zn), some functional expression of the proportion of blacks by town (b) and the Charles River dummy (chas) are consistently not included in the model. These variables would have thus been falsely identified as robust determinants for the willingness to pay for housing when using BMA or SSG-linear.

The SSG-smooth results further indicate that the two variables capturing the pupil-teacher ratio by town (ptratio) and the index of accessibility to radial highways (rad) only have a linear effect on the dependent variable. The covariates describing the average number of rooms per dwelling (rm), percentage of lower status of the population (lstat) and nitric oxides concentration (parts per 10 million; nox) are identified as having a linear effect with an additional nonlinear effect.

The DIC values indicate that SSG-smooth fits best, followed by BMA and SSG-linear. The DIC value for SSG-linear is considerably higher than for BMA. This might be due to the more complex model specification in SSG compared to BMA.

SSG-linear and SSG-smooth are also compared with respect to their posterior model size distributions. SSG-linear leads to an average model size of 11.1, whereas the additional inclusion of nonlinear terms leads to a slightly higher model size of 13.5. The corresponding visualization of the model size distributions is given in Figure 2 on the top left. For SSG-linear the maximum number of included terms is 13 and most of the models include either 11 or 12 out of the 13 terms. Similar to BMA, most of the potential covariates are included and hardly any variable selection is performed.

---

5 All variables are summarized with their names and a description in Table A1 in Appendix A.
| Variable | BMA | SSG-linear | SSG-smooth |
|----------|-----|------------|------------|
|          | Linear | Linear | Linear | Nonlin. | Linear | Linear | Linear | Nonlin. |
| dis      | 1.00 | −0.34 | 0.00 | 0.99 | −0.33 | 0.00 | 0.76 | −0.15 | 0.00 | 0.78 |
| lstat    | 1.00 | −0.41 | 0.00 | 1.00 | −0.41 | 0.00 | 1.00 | −0.47 | 0.00 | 0.97 |
| ptratio  | 1.00 | −0.22 | 0.00 | 1.00 | −0.22 | 0.00 | 0.97 | −0.16 | 0.00 | 0.13 |
| rm       | 1.00 | 0.29 | 1.00 | 1.00 | 0.30 | 1.00 | 1.00 | 0.27 | 1.00 | 1.00 |
| nox      | 1.00 | −0.22 | 0.00 | 0.97 | −0.21 | 0.00 | 0.66 | −0.13 | 0.01 | 0.80 |
| rad      | 1.00 | 0.28 | 1.00 | 0.94 | 0.23 | 1.00 | 0.77 | 0.22 | 1.00 | 0.12 |
| b        | 0.99 | 0.09 | 1.00 | 0.81 | 0.08 | 1.00 | 0.15 | 0.01 | 0.98 | 0.09 |
| tax      | 0.98 | −0.21 | 0.00 | 0.80 | −0.16 | 0.01 | 0.52 | −0.11 | 0.04 | 0.65 |
| zn       | 0.98 | 0.11 | 1.00 | 0.89 | 0.10 | 1.00 | 0.05 | 0.00 | 0.77 | 0.07 |
| crim     | 0.98 | −0.10 | 0.00 | 0.85 | −0.08 | 0.00 | 1.00 | −0.20 | 0.00 | 0.39 |
| chas     | 0.97 | 0.07 | 1.00 | 0.67 | 0.05 | 1.00 | 0.22 | 0.01 | 1.00 | – |
| indus    | 0.25 | 0.00 | 0.97 | 0.11 | −0.00 | 0.51 | 0.09 | −0.01 | 0.12 | 0.16 |
| age      | 0.24 | 0.00 | 0.98 | 0.11 | 0.00 | 0.50 | 0.05 | −0.00 | 0.39 | 0.07 |

For SSG-smooth, the posterior distribution of linear and nonlinear terms conditional on a given model size is shown in Figure 2 on the top right. The proportion of linear terms is shown in light grey. The proportion of nonlinear terms is split into the proportion where the corresponding linear term is also included (dark grey) or not (black). The widths of the bars are proportional to the posterior probability for models with this total number of terms. The plot indicates that even though the fitted models are larger for SSG-smooth, not all covariates are included. This is different from SSG-linear where nearly all covariates are always included and would have thus been identified as robust linear determinants for willingness to pay for housing. For SSG-smooth, only a subset of the covariates is included as having either a linear and/or nonlinear effect.

Figure 2 at the bottom visualizes the combined linear and nonlinear effect of the three covariates with the highest nonlinear PIPs. In the plot the 80% confidence bands and the median of their posterior distributions of the smooth effects are provided. The identified smooth effects are similar to those given in Smith and Kohn (1996, p. 337). For the average number of rooms per dwelling (rm) a strong price increase is only detected between approximately seven and eight rooms. Regarding the percentage of lower status of the population (lstat) there is a high decrease in price if the percentage is increased at a low level. This decrease is marginally diminishing, indicating that the price decrease is smaller if the percentage is already high. With respect to the nitric oxides concentration (nox) a price decrease occurs only after a certain level is reached which seems to be at approximately 0.6.

The results for the Boston housing data indicate that some covariates might be identified as having a robust linear effect using either BMA or SSG-linear, while their posterior
model inclusion probability turns out to be negligible once all covariates are included with potentially nonlinear effects. Thus, misleading results might be obtained if only linear terms are included in the model. The identification of robust effects indeed requires more flexible models to be considered.

V. Application to cross-country growth regressions

Two data sets for cross-country growth empirics are investigated: the data set used by Sala-i-Martin et al. (2004, referred to as the SDM data set) and the data set provided by Fernández et al. (2001, referred to as the FLS data set). These two data sets have already been thoroughly investigated with respect to the identification of robust linear effects. The aim of the presented analyses is to uncover robust nonlinear effects on growth. The design and specification used for analysing these two data sets are the same as for the Boston housing data set with details described in section ‘Design and specifications’ and online Appendix B.3.
The SDM data set

The data set by Sala-i-Martin et al. (2004) consists of data on 88 countries and 67 potential determinants of economic growth. Sala-i-Martin et al. (2004) used BMA techniques based on classical linear regression models to analyse this data set. They did not include any nonlinear terms.

Table 3 reports the results for BMA, SSG-linear and SSG-smooth. Only the results for those covariates are shown where at least one estimated PIP is larger than 0.1. Results are also shown for rather low PIPs in order to allow for a better comparison of results and assessment of the differences between the approaches, although variables with such low PIPs would not be identified as being robust determinants.

The results indicate that there are some differences between the estimated PIPs for the BMA and SSG-linear approach for this data set. The prior parameters for the slab distribution would need to be calibrated for BMA and SSG-linear in order to obtain a similar average posterior model size. However, even then still slight differences in the ordering of the variables according to their PIPs might occur due to the different shrinkage behaviour of the slab priors on the regression coefficients. Nonetheless, both linear approaches agree almost completely on the signs of the posterior means as indicated by the conditional positive sign certainties. The same robust positive or negative effects of the covariates on growth are identified.

SSG-smooth only identifies one additional variable as having a robust nonlinear effect on economic growth: population density (DENS60). The nonlinear term of this covariate has a posterior inclusion probability of 0.78 and is the most frequently included term in SSG-smooth, while its linear term has a PIP of less than 0.1 for BMA, SSG-linear as well as SSG-smooth. The importance of this covariate would thus have been neglected if only linear terms were included. Additional covariates which show a slight indication of a nonlinear effect are fraction GDP in mining (MINING) and malaria prevalence (MALFAL66). Both variables already have some support for inclusion as a linear term. For malaria prevalence, both the linear and the nonlinear effect have some posterior support indicating that the smooth effect estimated by SSG-smooth contains both components, while for fraction GDP in mining the linear effect has only a very small PIP once nonlinear effects are also included.

The DIC values indicate that SSG-smooth is slightly better fitting than SSG-linear, while overall BMA provides the best fit. Thus, based on the model fit criterion no clear preference for the model including nonlinear terms is discerned.

6 All variables are summarized with their names and a description in Table A2 in Appendix A.
7 See the discussion in online Appendix B.4 on the difference between BMA and SSG.
Regression under model uncertainty

TABLE 3
SDM data set analysed using BMA, SSG-linear and SSG-smooth

| Terms       | BMA | SSG-linear | SSG-smooth |
|-------------|-----|------------|------------|
| Variable    | Linear | Linear | Linear | Nonlin. | Linear | Linear | Linear | Nonlin. | Linear | Linear |
| EAST        | 0.82  | 0.31  | 1.00  | 0.63   | 0.25  | 1.00  | 0.69   | 0.26  | 1.00  | –      |
| P60         | 0.72  | 0.29  | 1.00  | 0.37   | 0.13  | 0.98  | 0.46   | 0.18  | 0.98  | 0.02   |
| IPRICE1     | 0.67  | -0.16 | 0.00  | 0.39   | -0.09 | 0.00  | 0.24   | -0.05 | 0.01  | 0.02   |
| GDPCH60L    | 0.60  | -0.25 | 0.00  | 0.17   | -0.06 | 0.04  | 0.05   | -0.01 | 0.10  | 0.03   |
| TROPICAR    | 0.48  | -0.17 | 0.00  | 0.27   | -0.08 | 0.02  | 0.31   | -0.10 | 0.01  | 0.04   |
| DENS65C     | 0.34  | 0.08  | 1.00  | 0.10   | 0.02  | 0.97  | 0.05   | 0.01  | 0.95  | 0.02   |
| MALFAL66    | 0.31  | -0.12 | 0.00  | 0.23   | -0.07 | 0.03  | 0.14   | -0.04 | 0.04  | 0.11   |
| LIFE060     | 0.23  | 0.11  | 1.00  | 0.17   | 0.07  | 0.94  | 0.08   | 0.03  | 0.90  | 0.09   |
| CONFUC      | 0.22  | 0.05  | 1.00  | 0.44   | 0.11  | 1.00  | 0.29   | 0.07  | 1.00  | 0.04   |
| SAFRICA     | 0.17  | -0.06 | 0.00  | 0.17   | -0.05 | 0.04  | 0.18   | -0.06 | 0.03  | –      |
| LAAM        | 0.16  | -0.04 | 0.02  | 0.09   | -0.02 | 0.06  | 0.06   | -0.01 | 0.08  | –      |
| MINING      | 0.15  | 0.02  | 1.00  | 0.08   | 0.01  | 0.96  | 0.02   | 0.00  | 0.89  | 0.12   |
| SPAIN       | 0.13  | -0.03 | 0.01  | 0.08   | -0.01 | 0.06  | 0.04   | -0.01 | 0.11  | –      |
| MUSLIM00    | 0.13  | 0.02  | 0.99  | 0.08   | 0.01  | 0.94  | 0.05   | 0.01  | 0.93  | 0.02   |
| GVR61       | 0.12  | -0.02 | 0.00  | 0.12   | -0.02 | 0.04  | 0.04   | -0.01 | 0.08  | 0.01   |
| BUDDHA      | 0.12  | 0.02  | 1.00  | 0.27   | 0.06  | 0.99  | 0.12   | 0.02  | 0.99  | 0.04   |
| YRSOPEN     | 0.12  | 0.03  | 1.00  | 0.20   | 0.05  | 0.97  | 0.13   | 0.04  | 0.96  | 0.02   |
| RERD        | 0.10  | -0.02 | 0.00  | 0.18   | -0.04 | 0.01  | 0.04   | -0.01 | 0.07  | 0.01   |
| DENS60      | 0.09  | 0.01  | 1.00  | 0.03   | 0.00  | 0.91  | 0.01   | 0.00  | 0.81  | 0.78   |
| TROPPPOP    | 0.07  | -0.01 | 0.01  | 0.12   | -0.03 | 0.07  | 0.04   | -0.00 | 0.27  | 0.03   |

**Notes:** nonlin: nonlinear; PIP: posterior inclusion probability; PM: posterior mean; PSC: posterior conditional positive sign certainty; DIC: deviance information criterion.

nonlinear effect, but that the covariate effects are either linear or nonlinear without a linear effect (i.e., fluctuating around a horizontal line).

The smooth effects of population density (DENS60), fraction GDP in mining (MINING), and malaria prevalence (MALFAL66) are visualized in Figure 3 at the bottom. For population density, clearly the effect is decreasing for small values, while this effect flattens out for average values. The increase at the end seems to be strongly influenced by a single observation, namely Botswana. A similar effect can be observed for fraction GDP in mining. For malaria prevalence the estimated smooth effect suggests that only high values of malaria prevalence are associated with lower growth.

These visualizations indicate that SSG-smooth is non-robust to outliers and care has to be taken when interpreting nonlinear effects. Nonlinear effects could be identified due to outliers, which do not conform with the general linear relationship between a covariate and the dependent variable. However, estimating only linear effects would also give misleading results. The estimated linear effect would be heavily influenced by the outlier. For example, the robust linear effect of population density (DENS60) identified for BMA and SSG-linear is positive as indicated by the PSC values which are equal to 1.00 and 0.91. However, this identified positive linear effect may be due only to the high leverage point Botswana. The
Figure 3. SDM data set. Top left: Distribution of model sizes for SSG-linear (black) and SSG-smooth (light grey). Top right: Distribution of linear terms (light grey), nonlinear terms with corresponding linear terms included (dark grey) and nonlinear terms without corresponding linear terms (black) given model size for SSG-smooth. The bar widths are proportional to the posterior probability of the model size. Bottom: Estimated robust smooth effects indicated by the 80% confidence bands and the median of the posterior distribution.

The visualization of the smooth effect suggests that for small values of population density there is actually a negative linear relationship between population density and economic growth.

Results for the SDM data set indicate that the population density variable has a low PIP for SSG-linear, but turns out to be the variable with the highest PIP when nonlinear terms are also included. The estimated nonlinear effect indicates that no global linear trend is discernible for this variable because (i) the effect of the variable differs for low and high values, and (ii) there exists also a high leverage point (i.e., Botswana) which strongly influences the estimated global linear trend if only linear terms are included. The non-robustness of linear regression thus can be alleviated by including nonlinear terms to identify outliers.

The FLS data set

The data set by Fernández et al. (2001) contains cross-country information on 72 countries and 41 potential growth determinants. The estimated nonlinear effect indicates that no global linear trend is discernible for this variable because (i) the effect of the variable differs for low and high values, and (ii) there exists also a high leverage point (i.e., Botswana) which strongly influences the estimated global linear trend if only linear terms are included. The non-robustness of linear regression thus can be alleviated by including nonlinear terms to identify outliers.

The FLS data set

The data set by Fernández et al. (2001) contains cross-country information on 72 countries and 41 potential growth determinants. Fernández et al. (2001) use BMA with only linear

---

8 All variables are summarized with their names and a description in Table A3 in Appendix A.

© 2019 The Authors. *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.

Electronic copy available at: https://ssrn.com/abstract=3615211
TABLE 4

| Terms       | BMA Linear | SSG-linear Linear | SSG-smooth Linear | Nonlin. |
|-------------|------------|-------------------|-------------------|---------|
| Variable    | PIP PM PSC | PIP PM PSC        | PIP PM PSC        | PIP     |
| GDP60       | 1.00 -0.79 0.00 | 0.73 -0.50 0.01 | 0.56 -0.36 0.01 | 0.05    |
| Confucian   | 0.99 0.29 1.00 | 0.83 0.25 1.00   | 0.74 0.22 1.00   | 0.03    |
| LifeExp     | 0.97 0.54 1.00 | 0.55 0.31 0.98   | 0.40 0.24 0.98   | 0.19    |
| EquipInv    | 0.92 0.25 1.00 | 0.74 0.31 0.99   | 0.71 0.31 0.99   | 0.07    |
| SubSahara   | 0.86 -0.34 0.00 | 0.60 -0.22 0.00 | 0.21 -0.07 0.02 | —       |
| Mining      | 0.74 0.12 1.00 | 0.19 0.04 0.97   | 0.10 0.02 0.96   | 0.14    |
| RuleofLaw   | 0.65 0.15 1.00 | 0.20 0.05 0.95   | 0.10 0.02 0.92   | 0.05    |
| NequipInv   | 0.63 0.10 1.00 | 0.18 0.03 0.96   | 0.18 0.04 0.98   | 0.19    |
| Muslim      | 0.58 0.12 1.00 | 0.19 0.04 0.94   | 0.12 0.02 0.95   | 0.07    |
| EcoOrg      | 0.58 0.09 1.00 | 0.17 0.03 0.97   | 0.11 0.02 0.97   | 0.03    |
| Hindu       | 0.55 -0.21 0.01 | 0.05 -0.00 0.18 | 0.02 -0.00 0.25 | 0.07    |
| Protestants | 0.53 -0.08 0.00 | 0.28 -0.06 0.02 | 0.16 -0.03 0.02 | 0.06    |
| LabForce    | 0.48 0.17 0.97 | 0.04 0.00 0.57   | 0.02 0.00 0.57   | 0.04    |
| LatAmerica  | 0.47 -0.12 0.03 | 0.26 -0.06 0.04 | 0.20 -0.05 0.03 | —       |
| BlMktPm     | 0.47 -0.05 0.00 | 0.06 -0.01 0.13 | 0.03 -0.00 0.23 | 0.10    |
| EthnoL      | 0.41 0.08 0.99 | 0.05 0.00 0.66   | 0.03 0.60 0.69   | 0.04    |
| YrsOpen     | 0.36 0.08 0.94 | 0.41 0.12 0.98   | 0.48 0.97 0.99   | 0.03    |
| Catholic    | 0.22 -0.01 0.39 | 0.09 -0.01 0.18 | 0.05 -0.01 0.15 | 0.17    |

DIC 99.22 172.86 167.03

Notes: nonlinear; PIP: posterior inclusion probability; PM: posterior mean; PSC: posterior conditional positive sign certainty; DIC: deviance information criterion.

terms to identify robust determinants of growth. Their results indicate that in total 18 covariates have a PIP of at least 10% and four variables have a PIP of at least 90% (namely, GDP level in 1960 (GDP60), fraction Confucian (Confucian), life expectancy (LifeExp), equipment investment (EquipInv)).

BMA, SSG-linear and SSG-smooth are used to fit the FLS data set. The results are shown in Table 4. Only those covariates are included where at least for one fitted model the PIP of the linear term is larger than 0.4 or the PIP of the nonlinear term is larger than 0.1. Again results for such low PIP values are only reported to enable a better comparison of the different results.

PIPs are clearly higher for BMA than for SSG-linear. To obtain a similar average posterior model size for both approaches the prior setting for the slab distributions would need to be calibrated. However, even if proportionally adjusting for the mean posterior model size some differences may remain for the PIPs obtained using the two approaches. This might be due to the different shrinkage behaviour of the two priors imposed on the regression coefficients. SSG adaptively shrinks each coefficient differently and also induces a stronger shrinkage on variables with large contribution to low-variance principal components (see online Appendix B.4). Furthermore, Hofmarcher et al. (2014) indicate that even if the same prior structure is used for the regression coefficients in a BMA analysis, the relative importance assigned to variables might change for increasing model...
size. They indicate that a variable which tends to be included for small model sizes might have a decreasing posterior weight for larger model sizes because several other variables are included for the larger model sizes essentially capturing the influence of this variable in a better way and replacing its inclusion.

Comparing SSG-linear and SSG-smooth indicates that most of the top ten variables identified for SSG-linear are also frequently included for SSG-smooth. Furthermore, for SSG-smooth the highest PIPs are observed only for linear terms. The nonlinear effect with the highest PIP is observed for non-equipment investment (NequipInv) and this variable has rank nine when effects are sorted by PIP. Furthermore, the DIC values reported are quite similar for SSG-linear and SSG-smooth indicating a slight preference for SSG-smooth. The DIC value for BMA is clearly smaller than for SSG-linear and SSG-smooth suggesting that a model including nonlinear terms is not required to achieve a good fit, but that BMA is preferable regarding model fit.

SSG-linear and SSG-smooth lead to rather similar posterior average model sizes (8.2 and 8.4). The posterior model size distributions for the two SSG approaches are shown in Figure 4 on the top left. For both the most frequent model size is eight. In Figure 4 on the
Regression under model uncertainty

It can be seen – similar to the SDM data set – that linear terms are included most often while nonlinear terms are only included if the corresponding linear term is not part of the model.

The smooth effects most frequently included in the model are visualized in Figure 4 at the bottom with their 80% confidence bands and the median of their posterior distributions. The smooth effects identified for non-equipment investment (NequipInv) and fraction Catholic (Catholic) provide a slight indication that extreme values for these covariates have a different effect than an average value. For life expectancy (LifeExp) the estimated robust smooth effect slightly indicates that the influence of the covariate shows diminishing returns for higher values.

For the FLS data set SSG-smooth validates and confirms the results obtained with SSG-linear and BMA. The results indicate that more complex functional relationships are not required to identify robust determinants and similar linear effects are discerned if the assumption of a constant linear effect is relaxed.

The results for the two growth data sets indicate that BMA might already be suitable for these data sets and the additional complexity of SSG is not required. Only slight nonlinear effects between the covariates and the dependent variable are detected or are even only identified because of outlying observations. Effects of outlying observations could also be captured within a standard BMA setting by either removing these observations or including dummy variables as covariates. However, these insights are only available if the more complex model is evaluated. Only then can the simpler model be safely selected and interpreted.

VI. Conclusions

In this paper the usefulness of SSG proposed by Scheipl et al. (2012) is indicated for economic regression analysis under variable and functional relationship uncertainty. SSG is a powerful tool to address two types of model uncertainty simultaneously: uncertainty about the explaining variables (variable uncertainty) and uncertainty about the functional form of the variables’ impact (linear vs. nonlinear effects). SSG enables researchers to identify the covariates with (i) robust linear effects and (ii) robust nonlinear effects on the outcome variable. The unknown nonlinear effects are modelled based on a flexible penalized spline approach. Relevant variables along with their functional relationships are identified based on stochastic search variable selection techniques.

The approach has several benefits: (i) misleading results are avoided, which might be obtained if nonlinear functional relationships between covariates and the dependent variable are neglected, (ii) results obtained using only linear terms can be validated by verifying that adding nonlinear effects does not change the identification of the robust linear effects, (iii) covariates without a linear, but only with a nonlinear effect, are not disregarded as important determinants for the dependent variable, (iv) differences in marginal effects can be identified which might have important policy implications, and (v) postprocessing tools available for BMA can be reused. SSG can also be employed in an exploratory way to identify nonlinear terms to be selectively included in a standard BMA application.

Finally, we want to emphasize that SSG-smooth can be easily applied using software available for the R environment for statistical computing and graphics (R Core Team, 2018).
and implemented in the package spikeSlabGAM (Scheipl, 2011). The results presented in this paper are obtained using the default setting for the prior parameters in the package. Thus, no extensive tuning of the hyper-parameters of the priors is required in order to arrive at useful solutions. Overall this suggests that any empirical economics analysis based on regression models and facing variable and functional relationship uncertainty should not rely on the assumption of ‘knowing’ a priori the functional relationship or imposing a simplified linear relationship. In contrast, robust determinants should be identified using suitable methods which account for variable and functional relationship uncertainty.

Appendix A

Tables A1–A3

TABLE A1

Variables and their description for the Boston housing data set

| Variable name | Description |
|---------------|-------------|
| crim          | Per capita crime rate by town |
| zn            | Proportion of residential land zoned for lots over 25,000 sq.ft |
| indus         | Proportion of non-retail business acres per town |
| chas          | Charles River dummy variable (= 1 if tract bounds river; 0 otherwise) |
| nox           | Nitric oxides concentration (parts per 10 million) |
| rm            | Average number of rooms per dwelling |
| age           | Proportion of owner-occupied units built prior to 1940 |
| dis           | Weighted distances to five Boston employment centres |
| rad           | Index of accessibility to radial highways |
| tax           | Full-value property-tax rate per USD 10,000 |
| pptratio      | Pupil-teacher ratio by town |
| b             | $1000(B - 0.63)^2$ where $B$ is the proportion of blacks by town |
| lstat         | Percentage of lower status of the population |
| medv          | Median value of owner-occupied homes in USD 1000’s |

TABLE A2

Variables and their description for the SDM data set

| Variable name | Description |
|---------------|-------------|
| ABSLATIT      | Absolute latitude |
| AIRDIST       | Air distance to big cities |
| AVELF         | Ethnolinguistic fractionalization |
| BRIT          | British colony |
| BUDDHA        | Fraction Buddhist |
| CATH00        | Fraction Catholic |
| CIV72         | Civil liberties |
| COLONY        | Colony dummy |

(continued)


TABLE A2

(Continued)

| Variable name | Description |
|---------------|-------------|
| CONFUC        | Fraction Confucian |
| DENS60        | Population density coastal 1960’s |
| DENS65C       | Population density 1960 |
| DENS65I       | Interior density |
| DPOP6090      | Population growth rate 1960–90 |
| EAST          | East Asian dummy |
| ECORG         | Capitalism |
| ENGFRA60      | English-speaking population |
| EUROPE        | European dummy |
| FERTLDC1      | Fertility in 1960’s |
| GDE1          | Defence spending share |
| GDPCH60L      | GDP 1960 (log) |
| GEEREC1       | Public education spending share in GDP in 1960’s |
| GGCFD3        | Government consumption share deflated with GDP prices |
| GOV1960M      | Nominal government GDP share 1960’s |
| GOVSH61       | Government share of GDP |
| GR6096        | Average growth rate of GDP per capita 1960–96 |
| GVR61         | Government consumption share 1960’s |
| H60           | Higher education in 1960 |
| HERF00        | Religious intensity |
| HINDU00       | Fraction Hindu |
| IPRICE1       | Investment price |
| LAAM          | Latin American dummy |
| LANDAREA      | Land area |
| LANDLOCK      | Landlocked country dummy |
| LHCPC         | Hydrocarbon deposits in 1993 |
| LIFE060       | Life expectancy in 1960 |
| LT100CR       | Fraction of land area near navigable water |
| MALFAL66      | Malaria prevalence in 1960’s |
| MINING        | Fraction GDP in mining |
| MUSLIM00      | Fraction Muslim |
| NEWSTATE      | Time of independence |
| OIL           | Oil-producing country dummy |
| OPENDEC1      | (Imports + exports)/GDP |
| ORTH00        | Fraction Orthodox |
| OTHFRAC       | Fraction speaking foreign language |
| P60           | Primary schooling 1960 |
| P6090         | Average speaking 1960–90 |
| POP1560       | Fraction population less than 15 |
| POP60         | Population in 1960 |
| POP6560       | Fraction population over 65 |
| PRI1970       | Primary exports in 1970 |
| PRIGHTS       | Political rights |
| PROT00        | Fraction Protestant |
| RERD          | Real exchange rate distortions |
| REVCOUN       | Revolution and coups |

(continued)
### TABLE A2

(Continued)

| Variable name | Description |
|----------------|-------------|
| SAfrica | African dummy |
| Scout | Outward orientation |
| Size60 | Size of the economy |
| Socialist | Socialist dummy |
| Spain | Spanish colony |
| SquI6090 | Square of inflation 1960–90 |
| TOT1DEC1 | Terms of trade growth in 1960’s |
| TOTIND | Terms of trade ranking |
| Tropicar | Fraction of tropical area |
| Troppop | Fraction population in tropics |
| Wartime | Fraction spent in war 1960–90 |
| Wartorn | War participation 1960–90 |
| Yrsopen | Years open |
| Ztropics | Tropical climate zone |

### TABLE A3

Variables and their description for the FLS data set

| Variable name | Description |
|----------------|-------------|
| Y | Economic growth 1960–92 as from the Penn World Tables Rev 6.0 |
| Abslat | Absolute latitude |
| Spanish | Spanish colony dummy |
| French | French colony dummy |
| Brit | British colony dummy |
| WarDummy | War dummy |
| LatAmerica | Latin America dummy |
| SubSahara | Sub-Sahara dummy |
| OutwarOr | Outward orientation |
| Area | Area surface |
| PrScEnroll | Primary school enrolment |
| LifeExp | Life expectancy |
| GDP60 | Initial GDP in 1960 |
| Mining | Fraction of GDP in mining |
| EcoOrg | Degree of capitalism |
| YrsOpen | Number of years having an open economy |
| Age | Age |
| Buddha | Fraction Buddhist |
| Catholic | Fraction Catholic |
| Confucian | Fraction Confucian |
| EthnoL | Ethnolinguistic fractionalization |
| Hindu | Fraction Hindu |
| Jewish | Fraction Jewish |
| Muslim | Fraction Muslim |
| PrExports | Primary exports 1970 |
| Protestants | Fraction Protestants |

(continued)
TABLE A3

| Variable name | Description                  |
|---------------|------------------------------|
| RuleofLaw     | Rule of law                  |
| Pog           | Population growth            |
| WorkPop       | Workers per inhabitant       |
| LabForce      | Size of labour force         |
| HighEnroll    | Higher education enrolment   |
| PublEdupct    | Public education share       |
| RevnCoup      | Revolutions and coups        |
| PolRights     | Political rights             |
| CivLib        | Civil liberties              |
| English       | Fraction speaking English    |
| Foreign       | Fraction speaking foreign language |
| RFEXDist      | Exchange rate distortions    |
| EquipInv      | Equipment investment         |
| NequipInv     | Non-equipment investment     |
| stdBMP        | Stand. dev. of black market premium |
| BlMktPm       | Black market premium         |

References

Amini, S. M. and Parmeter, C. F. (2012). ‘Comparison of model averaging techniques: assessing growth determinants’, *Journal of Applied Econometrics*, 27, 870–876.

Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2015). ‘Dirichlet-Laplace priors for optimal shrinkage’, *Journal of the American Statistical Association*, 110, 1479–1490.

Brock, W. and Durlauf, S. (2001). ‘Growth empirics and reality’, *World Bank Economic Review*, 15, 229–272.

Ciccone, A. and Jarociński, M. (2010). ‘Determinants of economic growth: will data tell?’, *American Economic Journal: Macroeconomics*, 2, 222–246.

Crespo Cuaresma, J. (2010). ‘How different is Africa? A comment on Masanjala and Papageorgiou’, *Journal of Applied Econometrics*, 26, 1041–1047.

Delgado, M. S., Henderson, D. J. and Parmeter, C. F. (2014). ‘Does education matter for economic growth?’, *Oxford Bulletin of Economics and Statistics*, 76, 334–359.

Eicher, T. S., Henn, C. and Papageorgiou, C. (2012). ‘Trade creation and diversion revisited: accounting for model uncertainty and natural trading partner effects’, *Journal of Applied Econometrics*, 27, 296–321.

Eilers, P. H. C., Marx, B. D. (1996). ‘Flexible smoothing with B-splines and penalties’, *Statistical Science*, 11, 89–102.

Fahrmeir, L., Kneib, T. and Konrath, S. (2010). ‘Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection’, *Statistics and Computing*, 20, 203–219.

Fernández, C., Ley, E. and Steel, M. F. J. (2001). ‘Model uncertainty in cross-country growth regressions’, *Journal of Applied Econometrics*, 16, 563–576.

Harrison Jr., D. and Rubinfeld, D. L. (1978). ‘Hedonic housing prices and the demand for clean air’, *Journal of Environmental Economics and Management*, 5, 81–102.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*, Cambridge University Press, Cambridge.
Henderson, D. J. and Parmeter, C. F. (2016). ‘Model averaging over nonparametric estimators’, in González-Rivera G., Hill R. C. and Lee T. H. (eds.), Essays in Honor of Aman Ullah, volume 36 of Advances in Econometrics, Bingley, UK: Emerald Group Publishing Limited, pp. 539–560.

Henderson, D. J., Papageorgiou, C. and Parmeter, C. F. (2011). ‘Growth empirics without parameters’, The Economic Journal, 122, 125–154.

Hofmarcher, P., Kerbl, S., Grün, B., Sigmund, M. and Hornik, K. (2014). ‘Model uncertainty and aggregated default probabilities: new evidence from Austria’, Applied Economics, 46, 871–879.

Hofmarcher, P., Crespo Cuáresma, J., Grün B., Humer S. and Moser, M. (2018). ‘Bivariate jointness measures in Bayesian model averaging: solving the conundrum’, Journal of Macroeconomics, 57, 150–165.

Ishwaran, H. and Rao, J. S. (2005). ‘Spike and slab variable selection: frequentist and Bayesian strategies’, The Annals of Statistics, 33, 730–773.

Koop, G. and Tobias, J. L. (2006). ‘Semiparametric Bayesian inference in smooth coefficient models’, Journal of Econometrics, 134, 283–315.

Lang, S. and Brezger, A. (2004). ‘Bayesian P-splines’, Journal of Computational and Graphical Statistics, 13, 183–212.

Ley, E. and Steel, M. F. J. (2012). ‘Mixtures of g-priors for Bayesian model averaging with economic applications’, Journal of Econometrics, 171, 251–266.

Li, Q. and Racine, J. S. (2007). Nonparametric Econometrics: Theory and Practice, Princeton University Press, Princeton, NJ.

Malsiner-Walli, G. and Wagner, H. (2011). ‘Comparing spike and slab priors for Bayesian variable selection’, Austrian Journal of Statistics, 40, 241–264.

Mitchell, T. J. and Beauchamp, J. J. (1988). ‘Bayesian variable selection in linear regression’, Journal of the American Statistical Association, 83, 1023–1032.

Montgomery, J. M. and Nyhan, B. (2010). ‘Bayesian model averaging: theoretical developments and practical applications’, Political Analysis, 18, 245–270.

O’Hara, R. B. and Sillanpää, M. J. (2009). ‘A review of Bayesian variable selection methods: what, how and which’, Bayesian Analysis, 4, 85–118.

Park, T. and Casella, G. (2008). ‘The Bayesian lasso’, Journal of the American Statistical Association, 103, 681–686.

R Core Team (2018). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Rockey, J. and Temple, J. (2016). ‘Growth econometrics for agnostics and true believers’, European Economic Review, 81, 86–102.

Sala-i-Martin, X., Doppelhofer, G. and Miller, R. I. (2004). ‘Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach’, American Economic Review, 94, 813–835.

Scheipl, F. (2011). ‘spikeSlabGAM: Bayesian variable selection, model choice and regularization for generalized additive mixed models in R’, Journal of Statistical Software, 43, 1–24.

Smith, M. and Kohn, R. (1996). ‘Nonparametric regression using Bayesian variable selection’, Journal of Econometrics, 75, 317–343.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. (2002). ‘Bayesian measures of model complexity and fit’, Journal of the Royal Statistical Society, Series B, 64, 583–639.

Tibshirani, R. (1996). ‘Regression shrinkage and selection via the lasso’, Journal of the Royal Statistical Society, Series B, 58, 267–288.

Wood, S. N. (2006). Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC Press, Boca Raton, FL.

Zeugner, S. and Feldkircher, M. (2015). ‘Bayesian model averaging employing fixed and flexible priors: the BMS package for R’, Journal of Statistical Software, 68, 1–37.
Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix B. Methodological details.