Last Night a Shrinkage Saved My Life: Economic Growth, Model Uncertainty and Correlated Regressors

Hofmarcher, Paul; Crespo Cuaresma, Jesus; Grün, Bettina; Hornik, Kurt

Published in:
Journal of Forecasting

DOI:
10.1002/for.2328

Published: 01/01/2015

Document Version
Peer reviewed version

Citation for published version (APA):
Hofmarcher, P., Crespo Cuaresma, J., Grün, B., & Hornik, K. (2015). Last Night a Shrinkage Saved My Life: Economic Growth, Model Uncertainty and Correlated Regressors. *Journal of Forecasting*, *34*(2), 133 - 144. https://doi.org/10.1002/for.2328

Last Night a Shrinkage Saved My Life: Economic Growth, Model Uncertainty and Correlated Regressors

Paul Hofmarcher*
WU

Jesús Crespo Cuaresma†
WU, IIASA, WIC and *WIFO*

Bettina Grün‡
JKU

Kurt Hornik§
WU

Abstract

We compare the predictive ability of Bayesian methods which deal simultaneously with model uncertainty and correlated regressors in the framework of cross-country growth regressions. In particular, we assess methods with *spike and slab* priors combined with different prior specifications for the slope parameters in the slab. Our results indicate that moving away from Gaussian g-priors towards Bayesian ridge, LASSO or elastic net specifications has clear advantages for prediction when dealing with datasets of (potentially highly) correlated regressors, a pervasive characteristic of the data used hitherto in the econometric literature.

*Department of Economics, Vienna University of Economics and Business (WU). Welthandelsplatz 1, 1020 Vienna, Austria. paul.hofmarcher@wu.ac.at

†Department of Economics, Vienna University of Economics and Business (WU). Welthandelsplatz 1, 1020 Vienna, Austria; World Population Program, International Institute of Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital (WIC) and Austrian Institute for Economic Research (WIFO). jcrespo@wu.ac.at.

‡Department of Applied Statistics, Johannes Kepler University Linz. Altenbergerstraße 69, 4040 Linz, Austria. bettina.gruen@jku.at

§Department of Finance, Accounting and Statistics, Vienna University of Economics and Business (WU). Welthandelsplatz 1, 1020 Vienna, Austria. kurt.hornik@wu.ac.at

1 Introduction

Inference under model uncertainty is a pervasive problem of empirical applications in economics. In particular, assessing empirically the robustness of economic growth determinants under model uncertainty is a subject which has spawned many econometric studies in the last decade. Fernández et al. (2001b), Sala-i-Martin et al. (2004), Crespo Cuaresma and Doppelhofer (2007), Ley and Steel (2007), Doppelhofer and Weeks (2009), Ley and Steel (2009), Durlauf et al. (2008), Eicher et al. (2011) or Ley and Steel (2012) are some examples of studies which apply methods based on Bayesian model averaging to account for uncertainty in the specification of econometric models aimed at explaining differences in long-run economic growth across countries.

Most of the existing methods used in this branch of the literature do not assess explicitly the potential problem of multicollinearity among the set of potential explanatory variables. Although some *ad hoc* dilution priors have been proposed in the literature to account for related regressors (see for example Durlauf et al. (2008), who puts forward the use of the correlation matrix of model-specific regressors to adjust model priors based on the idea of dilution priors put forward by George (1999)), a systematic assessment of the issue is hitherto missing. A notable exception is the recent work by Korobilis (2013), which evaluates Bayesian variable selection in regressions with a large number of potentially highly correlated predictors using clustering methods by means of the prior proposed by Dunson et al. (2008).[1] In this paper we evaluate a large number of available Bayesian methods aimed at dealing with the problem of model uncertainty in the presence of correlated regressors. In addition to standard Gaussian g-priors, we also assess prior structures that have been proposed in the framework of Bayesian bridge regression, which allow us to deal explicitly with the problem of correlated explanatory variables by shrinking coefficients. Together with the most prominent cases of the bridge regression class (ridge regression and LASSO), we also include Bayesian elastic net specifications in our analysis. This specification nests both ridge regression and the LASSO estimator as special cases. The set of shrinkage methods studied is embedded in specifications which use *spike and slab* priors (see for instance Mitchell and Beauchamp (1988) or George and McCulloch (1993)). This allows us to deal with variable selection or model averaging and account for the correlation structure of regressors simultaneously in a systematic way. Furthermore, the use of a *spike and slab* prior allows us to include explicitly prior information concerning model size or the relative importance of covariates in the specification in a straightforward manner.[2]

In order to assess quantitatively the importance of accounting for correlated regressors under model uncertainty in the setting of cross-country growth regressions, we provide a thorough comparison of Bayesian shrinkage methods in terms of their out-of-sample predictive accuracy. We evaluate the relative predictive ability of the different methods proposed in the literature making use of the dataset by Sala-i-Martin et al. (2004), which comprises information on income per capita growth for the period 1960–1996 and 67 potential growth determinants for a broad cross-section of countries. Schneider and Wagner (2012) apply frequentist adaptive LASSO methods to the dataset and find a substantial degree of similarity with the results in Sala-i-Martin et al. (2004), although some variables (*Population Coastal Density* or *Life Expectancy*, for instance) which Sala-i-Martin et al. (2004) tagged as robust do not appear to be important according to the results using the shrinkage method. We show that the use of shrinkage methods that are designed to deal

[1] Related regressors in Bayesian model averaging have been assessed more deeply in the framework of interaction terms and polynomial specifications (see Chipman (1996) for a general presentation and Crespo Cuaresma (2011) for a recent application to economic growth).

[2] Related Bayesian stochastic search methods in the framework of multivariate time series models can be found in George et al. (2008) and Jochmann et al. (2010), for instance.

with correlated regressors leads to some important changes in the results of the in-sample robustness analysis as compared to the existing literature. On the one hand, the posterior model probability is more spread across specifications. On the other hand, as in Schneider and Wagner (2012), compared to Sala-i-Martin et al. (2004) the *Life Expectancy* variable strongly reduces its importance in our results, but variables like *Malaria Prevalence* and *Fraction Confucian* appear as more robust growth determinants. Our out-of-sample prediction exercise confirms the superiority of shrinkage methods such as those implied by Bayesian ridge, LASSO and elastic nets over standard g-priors, both for the case of fixed g-priors and for specifications which propose hyperpriors for the g-parameter (see Liang et al. (2008), Feldkircher and Zeugner (2009) or Ley and Steel (2012)).

The paper is structured as follows. Section 2 presents the basic specification and the set of priors that are evaluated. Section 3 presents in-sample results based on the data by Sala-i-Martin et al. (2004) which show the differences in the nature of robust determinants of economic growth depending on the prior used. Section 4 presents the analysis of out-of-sample predictive ability and Section 5 concludes.

2 Economic Growth, Model Uncertainty and Correlated Regressors

2.1 Linear Regression, Model Uncertainty and the *Spike and Slab* Prior

Assume that a group of K explanatory variables $\{x_1, \ldots, x_K\}$ are proposed as potential determinants of economic growth (y) differences across countries in the framework of linear regression models. Let the specification where all K variables are included in the model be given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

where $\mathbf{y}$ is a vector containing the N observations of income per capita growth, $\mathbf{X}$ is the $N \times K$ design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1 \ldots \beta_K)'$ denotes the parameter vector of interest and it is assumed that $\mathbf{u} \sim \mathbf{N}(0, \sigma^2 \mathbf{I}_N)$.

In order to integrate explicitly the model uncertainty dimension into the estimation procedure, we impose a prior over the elements of the $\boldsymbol{\beta}$ vector which corresponds to a *spike and slab* mixture such as that put forward by Mitchell and Beauchamp (1988) (see also George and McCulloch (1993) and George and McCulloch (1997)). We assign a prior to each single coefficient β_j which is a mixture of a point mass at zero and some probability distribution for β_j.[3] That implies that the prior on β_j is given by

$$p(\beta_j|\gamma_j, \sigma^2) \sim (1 - \gamma_j) I_0 + \gamma_j \pi(\beta_j|\sigma^2), \tag{2}$$

where the main aim of our study is to compare the predictive ability of models which differ in terms of the specification of $\pi(\beta_j|\sigma^2)$. A Bernoulli prior is assumed on γ_j, so that $\gamma_j \sim \text{Be}(\underline{\gamma})$. We can elicit the prior by setting $\underline{\gamma} = \bar{k}/K$, where $\bar{k}$ can be interpreted as the expected value of the prior over model size. The standard specification is nested in this setting and corresponds to imposing $\bar{k} = K$. The posterior distribution of γ_j, $p(\gamma_j|y)$ can be interpreted based on the concept of posterior inclusion probability (PIP), which is widely used in the modern literature on Bayesian model averaging as an indicator of robustness of covariates to model uncertainty (see for example Fernández et al. (2001b), Sala-i-Martin et al. (2004) or Ley and Steel (2009) for empirical applications related to

[3]The point mass at zero is also sometimes replaced by a mean zero normal distribution with a very low variance (see e.g. George and McCulloch (1993)).

economic growth).

Ley and Steel (2009), following Brown et al. (1998), propose to robustify the choice of a prior variable inclusion probability (and thus, of a prior expected model size) by imposing a hyperprior on $\underline{\gamma}$, so that $\underline{\gamma} \sim \text{Beta}(a, b)$. They show that inference based on such a hyperprior over the prior inclusion probability makes on the one hand inference more robust to the choice of a prior expected model size and on the other hand it improves the out-of-sample predictive ability of model averaging techniques. We also follow this approach in our empirical application.

We proceed by describing different approaches to specifying $\pi(\beta_j|\sigma^2)$ which have been proposed in the literature and concentrate on methods which deal with correlated regressors.

2.2 Zellner's *g*-type priors

Following Zellner (1986), so-called g-priors have become a usual choice for models such as those given by (1) and (2). For a model with design matrix $\mathbf{X}_s$, the g-prior implies specifying $\pi(\beta_j|\sigma^2)$ as the corresponding element of a multivariate normal distribution with zero expected value and a variance-covariance matrix given by $\sigma^2(g\mathbf{X}_s{}'\mathbf{X}_s)^{-1}$. Such a prior has the advantage that it only requires the elicitation of one parameter (g) and the resulting marginal likelihood can easily be evaluated analytically (see for example Fernández et al. (2001a)). Standard choices for the elicitation of g proposed in the literature include:

- $g = N$ (*unit information prior*, UIP, see Kass and Wasserman (1995) and Kass and Raftery (1995)), which corresponds to choosing a prior where the amount of information about the parameter equals that contained in one observation and leads to Bayes factors that behave like the Bayesian Information Criterion (BIC),
- $g = K^2$ (*risk inflation criterion*, RIC, see Foster and George (1994)),
- $g = \max(N, K^2)$ (*benchmark g-prior*, BRIC, see Fernández et al. (2001a)),
- $g_s = \arg\max_g p(y|\mathbf{X}_s, g)$ (*empirical Bayes-local prior*, see George and Foster (2000), Hansen and Yu (2001) and Liang et al. (2008)), which implies estimating a different g parameter for each model based on the corresponding marginal likelihood,
- setting a hyperprior, which is often defined on the shrinkage factor, $g/(1+g)$ (see Liang et al. (2008), Feldkircher and Zeugner (2009) and Ley and Steel (2012)).

2.3 Ridge Regression, LASSO and the Elastic Net

Assuming that $N > K$, the standard ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ in (1), $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, will have unsatisfactory features if the design matrix is ill-conditioned, that is, if the explanatory variables are highly correlated. In particular, notice that $E((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \sigma^2 \sum_{j=1}^{K} \lambda_j^{-1}$, where $\{\lambda_1, \ldots, \lambda_K\}$ are the eigenvalues of $\mathbf{X}'\mathbf{X}$.[4] If multicollinearity among our regressors is present, at least one of the eigenvalues will be close to zero, inflating the variance of the OLS estimator.

Bridge regression methods have been proposed in order to deal with this problem. In a frequentist setting, the bridge regression estimate is obtained by minimizing the residual sum of squares subject to the constraint $\sum_{j=1}^{K} |\beta_j|^\gamma < t$ for constants t and $\gamma \geq 1$. The

[4] See, e.g. Poirier (1995), page 582.

regression coefficients are thus obtained as

$$\hat{\boldsymbol{\beta}}_{\text{bridge}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{K} |\beta_j|^{\gamma} \right\}. \tag{3}$$

The Lagrangian parameter $\lambda \geq 0$ can be interpreted as a shrinkage weight and γ defines the differential shrinkage of parameters. Prominent estimators derived from Equation (3) are the *ridge regression* (Hoerl and Kennard (1970)) estimator, with $\gamma = 2$, and the *least absolute shrinkage and selection operator* (LASSO) estimator (see Tibshirani (1996)), for which $\gamma = 1$. However, while the penalty in (3) shrinks parameters for $\gamma = 2$, it does not necessarily set them to zero. The form of the shrinkage in the LASSO estimator allows for corner solutions with some elements of $\boldsymbol{\beta}$ equal to zero. In this sense, the LASSO estimator acts at least partly as a model selection device.[5]

From a Bayesian point of view, the ridge and LASSO estimators appear as posterior mode estimators under particular prior settings (see for example Hans (2009) and Park and Casella (2008)). Both estimators can be obtained in the framework of a Bayesian hierarchical model where the distribution of the regression coefficients is given by a scale mixture of normal distributions with mixing over $\boldsymbol{\tau}^2$. Conditional on $\boldsymbol{\tau}^2$ the prior distribution of the regression cofficients is given by

$$\boldsymbol{\beta}|\boldsymbol{\tau}^2, \sigma^2 \sim \mathbf{N}(0, \sigma^2 \mathbf{W}_{\boldsymbol{\tau}}), \tag{4}$$

with $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau_K^2)$ and $\mathbf{W}_{\boldsymbol{\tau}} = \text{diag}\{\tau_1^2 \ldots \tau_K^2\}$. The standard improper prior over the error variance is used,

$$p(\sigma^2) \propto 1/\sigma^2, \tag{5}$$

and the LASSO estimator is obtained by assigning an independent exponential distribution as prior for each τ_j^2. The ridge regression estimator, on the other hand, is obtained by a degenerate mixture where $\boldsymbol{\tau}^2$ has a fixed value.

From a frequentist perspective, the elastic net uses a convex combination of the penalties implied by the ridge and LASSO regression and therefore obtains the estimator as

$$\hat{\boldsymbol{\beta}}_{\text{enet}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \sum_{j=1}^{K} \left(\lambda_1 |\beta_j| + \lambda_2 \beta_j^2 \right) \right\}. \tag{6}$$

The elastic net combines thus the characteristics of the ridge regression and the LASSO. Li and Lin (2010) and Bornn et al. (2010) present a Bayesian framework to estimate elastic nets. Following Li and Lin (2010), the following prior is assigned to the parameters of the model

$$\boldsymbol{\beta}|\sigma^2 \sim \exp\left\{ -\frac{1}{2\sigma^2} \left[\lambda_1 \sum_{j=1}^{K} |\beta_j| + \lambda_2 \sum_{j=1}^{K} \beta_j^2 \right] \right\}. \tag{7}$$

This prior over $\boldsymbol{\beta}$ conditional on σ^2, combined with (5) and the fact that $\mathbf{y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_N)$, allows for the use of a Gibbs sampler to estimate the corresponding posterior distributions.

[5]When it comes to optimization, there is still some reluctance to adopt L_1 methods of estimation, although Portnoy and Koenker (1997) demonstrate that L_1 regression ($\gamma = 1$) can be made competitive with L_2 regression ($\gamma = 2$) in terms of computational speed.

Posterior distributions for the parameters of interest can be obtained after noting that, as for the case of the LASSO and ridge regression, conditional on σ^2, the distribution of β_j can be treated as a scale mixture of normals. In the case of the Bayesian elastic net, as shown in Li and Lin (2010), the mixing distribution is given by a truncated Gamma distribution.

To the extent that parameter estimates in the Bayesian elastic net framework are shrunk to zero, the model embodies to a certain degree a variable selection mechanism. Given the logic behind shrinkage models, such a mechanism takes explicitly into account the potential effect of multicollinearity. The existing studies on Bayesian elastic nets propose carrying out variable selection through *ad hoc* approaches based on the posterior distribution of the individual elements in $\boldsymbol{\beta}$. Li and Lin (2010) propose using the *credible interval* and *scaled neighborhood* criteria. Using the former, a variable x_j is excluded if the credible interval of its corresponding parameter covers zero. The latter one considers the posterior probability contained in $[-\sqrt{\text{var}(\beta_j|y)}, \sqrt{\text{var}(\beta_j|y)}]$ and a variable is excluded if this posterior probability exceeds a certain ad hoc probability threshold.

It should be noticed that the use of the spike and slab prior has several advantages as compared to relying exclusively on the variable selection method embodied in the shrinkage strategy of the elastic net. By controlling the prior expected model size through the elicitation of $\underline{\gamma}$, we are able to exploit additional prior information concerning our beliefs about the number of variables which should be included in the specification. In applications related to model averaging and model comparison in the framework of cross-country growth regressions, for example, models with a very large number of covariates tend to be considered "less probable" *a priori* than models with a relatively small size. In terms of model comparison, the inclusion of such a prior over the model space implies that, in addition to the penalty on model size embodied in the Bayes factor, very large models may be further penalized using a prior model probability which depends on the number of covariates included in the specification. Specifications with spike and slab priors expanded with an additional hierarchical shrinkage such as that put forward here are used by Yuan and Lin (2005) in a setting that, under an appropriate parametrization, nests the models described above (see Kyung et al. (2010)).

The setting presented implies that inference on the parameters of our model is subject to two types of shrinkage mechanisms. On the one hand, the potential multicollinearity present in the set of covariates is explicitly taken into account by the automatic shrinkage imposed by the LASSO, ridge or elastic net shrinkage. On the other hand, the relative *a priori* importance of each variable as a determinant of y (or the relative prior belief that the size of the model is "reasonable") determines a second type of shrinkage which is implemented through the spike and slab structure given by (2). The full model can be estimated in a straightforward manner by integrating the Gibbs sampling procedure proposed by Li and Lin (2010) into the structure of the Gibbs sampler used to estimate linear models with spike and slab priors (as described in e.g. Mitchell and Beauchamp (1988)).

3 Correlated Regressors and the Determinants of Economic Growth: An In-Sample Exploration

Sala-i-Martin et al. (2004) study the robustness of economic growth determinants to model uncertainty using a dataset for 88 countries comprising data on income per capita growth over the period 1960–1996 as well as 67 variables which have been proposed as potential determinants of income growth in the literature. The dataset has become a workhorse to apply econometric methods related to model uncertainty and model averaging (see Ley and Steel (2007), Doppelhofer and Weeks (2009), Ley and Steel (2009) or Eicher et al. (2011), just to name a few, for recent papers where new techniques related to Bayesian model averaging are applied to these data).[6] The average absolute correlation between the variables is only 0.212. However we observe groups of highly correlated variables, such as for example *Political Rights*, *Fraction Population Less than 15*, *Fraction Population Over 65*, *European Dummy*, *Fertility Rates in 1960s* and *Population Growth Rate 1960–90* with an average absolute correlation of 0.794.

Before assessing the differences in out-of-sample predictive ability across the set of priors described in Section 2, we aim at comparing the in-sample results concerning the robustness of economic growth determinants which emerge for different shrinkage methods. In particular, we exemplify the differences by comparing the parameter estimates obtained by applying standard Bayesian methods with a spike and slab prior (using a beta-binomial hyperprior) and a benchmark g-prior (as in Ley and Steel (2009)) with those obtained using the parameter priors that lead to a Bayesian elastic net model. We use this exercise to evaluate how two fundamentally different priors over the slope coefficients in the economic growth regression affect our inference about the relative importance of the covariates proposed in the literature as determinants of income growth. Given the flexibility of the elastic net specification to replicate both ridge and LASSO estimators, we choose it as a representative example of the class of methods that go beyond the shrinkage implied by g-priors. For the estimation of the elastic net model using the data in Sala-i-Martin et al. (2004), we employ the following uninformative priors. We use a Beta$(1, 1)$ prior on $\underline{\gamma}$, and reparametrize λ_1 and λ_2 as $\alpha\lambda$ and $(1-\alpha)\lambda$, respectively, imposing the same prior structure as for $\underline{\gamma}$ on α. We introduce a hyperprior on λ, so that $\lambda^2 \sim$ Gamma$(0.1, 0.1)$.[7] The precision of the error term $\mathbf{u}$, $1/\sigma^2$, is assumed to follow a Gamma$(0.001, 0.001)$ and each τ_j is drawn from a $[1, \infty)$ truncated Gamma distribution with a shape value of 0.5. The Gibbs sampler is implemented by running four parallel Markov chains, each initialized with a different seed. One million iterations of the sampler were performed, whereby only every tenth value was used for posterior estimation. Convergence diagnostic indicated satisfactory convergence and the results presented are based on averages over the individual Markov chains.

Table 1 compares the results obtained using the BRIC g-prior with those obtained from estimating the Bayesian elastic net with spike and slab priors on the inclusion of the variables.[8] The standard estimation setting, labelled BRIC in Table 1, corresponds to that proposed by Ley and Steel (2009) for the same dataset. The first column presents the ranking in terms of posterior inclusion probabilities (PIP, the mean of the posterior distribution of γ_j, which can be loosely interpreted as the probability that the variable is included in the true model) implied by the results of the standard BRIC prior. In the following three columns the PIP, as well as the mean and standard deviation of the posterior distribution of each parameter are shown for the standard BRIC setting, together with those obtained using the Bayesian elastic net. The PIP of the results using the BRIC prior and those of the elastic net have a correlation of 0.81, although some strong differences can be observed when comparing the relative importance of the variables in

[6]The dataset can be obtained from Gernot Doppelhofer's homepage at `http://www.nhh.no/Default.aspx?ID=3075.`

[7]We depart here from the proposal by Li and Lin (2010), who put forward to use an empirical Bayes prior for λ_1 and λ_2. Our approach is based on Park and Casella (2008), and is also proposed by Li and Lin (2010) as an alternative to empirical Bayes.

[8]All the computations within this work are done by using R, a language and environment for statistical computing (R Core Team (2012)) and its extension packages rjags, coda and BMS. Codes are available from the authors upon request.

the dataset. The mean of the posterior distribution of γ_j in the elastic net specification is 0.139, corresponding to a mean of the posterior model size distribution of approximately 9.3, a result which is in line with the results presented for the same dataset by Ley and Steel (2009). The correlation between posterior means of the parameter estimates is 0.61.

The shrinkage implied by the Bayesian elastic net has effects on the nature of the robust determinants of economic growth implied by the results in Table 1. The posterior model probability appears much more spread across specifications than in the standard BRIC case, where it is concentrated on few models containing an *East Asian Dummy* and the variable *Malaria Prevalence*. Some variables, such as *Ethnolinguistic Fractionalization*, level and duration of *Openness*, *Initial Income* or *Fraction Confucian* strongly increase their PIPs under the Bayesian elastic net setting. On the other hand, *Life Expectancy* and *Investment Price* decrease their relative importance radically when the Bayesian elastic net is used instead of the BRIC g-prior, although their posterior inclusion probabilities are not strongly affected.[9]

For approaches resulting in similiar posterior model sizes, the differences in results between those methods can be traced back to the way that correlated regressors are dealt with. A standard measure for the degree of collinearity among the variables in a given model is given by the determinant $|\mathbf{R}_s|$ of the correlation matrix of regressors, a measure proposed by George (1999) as a building block of dilution priors over the model space. This determinant equals one if the columns of $\mathbf{X}_s$ are orthogonal and zero for perfectly collinear columns in $\mathbf{X}_s$. As the standard approach with the setting in Ley and Steel (2009) results in smaller posterior model sizes those in the elastic net approach, we depart from the hyper-prior over covariate inclusion and employ a uniform model prior instead. This results in comparable posterior model sizes. In particular, in this setting, we observe an average posterior model size of 8.9 for the standard approach and 9.2 for the elastic net approach. We compute the determinant of the correlation matrix of regressors for all models visited by the Markov chain for each of the two methods evaluated and the histograms of $|\mathbf{R}_s|$ are shown in Figure 1. The standard approach has an average determinant of the correlation matrix of regressors of 0.092, while for the Bayesian elastic net the mean determinant is 0.179, nearly twice as large. A larger number of models with very small regressor correlation determinants are visited in the standard BMA approach, while for the Bayesian elastic net method models with high values for the determinants (above 0.7) are also visited.

Interpreting our results in the context of model averaging techniques, these indicate that methods dealing with correlated regressors using hierarchical priors of the type employed in the Bayesian elastic net lead to averaging over models whose explanatory variables are on average less collinear. This implies that variables with a high correlation to other important variables but with a small effect on the dependent variable tend indeed to be omitted due to the regularization effect implied by the shrinkage of the Bayesian elastic net.

4 Economic Growth, Shrinkage Priors and Out-Of-Sample Predictive Ability

In order to evaluate the relative advantages of using different shrinkage priors in the context of cross-country growth regressions, we compare the forecasting ability of the

[9]The results obtained using Bayesian LASSO and Bayesian ridge regression do not differ qualitatively from those presented for the Bayesian elastic net. They are available from the authors upon request.

#	Description	Name	LS, 2009 PIP	LS, 2009 PM	LS, 2009 PSD	Bayesian elastic net PIP	Bayesian elastic net PM	Bayesian elastic net PSD
1	East Asian Dummy	EAST	0.992	0.030	0.005	0.976	0.024	0.007
2	Malaria Prevalence	MALFAL66	0.863	−0.020	0.009	0.657	−0.015	0.006
3	Primary Schooling Enrollment	P60	0.125	0.003	0.007	0.590	0.017	0.008
4	Fraction of Tropical Area	TROPICAR	0.101	−0.001	0.004	0.367	−0.010	0.005
7	Years Open 1950-94	YRSOPEN	0.030	0.000	0.003	0.364	0.012	0.006
12	Fraction Confucian	CONFUC	0.013	0.001	0.008	0.229	0.018	0.019
6	Spanish Colony Dummy	SPAIN	0.034	0.000	0.002	0.215	−0.008	0.005
8	Government Consumption Share	GVR61	0.027	−0.002	0.010	0.208	−0.017	0.019
18	Initial Income (Log GDP in 1960)	GDPCH60L	0.008	0.000	0.001	0.205	−0.006	0.003
13	Ethnolinguistic Fractionalization	AVELF	0.012	0.000	0.002	0.202	−0.009	0.007
21	(Imports + Exports)/GDP	OPENDEC1	0.006	0.000	0.001	0.200	0.007	0.005
10	Primary Exports in 1970	PRIEXP70	0.018	0.000	0.002	0.191	−0.009	0.007
11	Fraction Population In Tropics	TROPPOP	0.014	0.000	0.002	0.190	−0.009	0.007
24	Sub-Saharan Africa Dummy	SAFRICA	0.004	0.000	0.001	0.185	−0.008	0.006
30	Fraction GDP in Mining	MINING	0.002	0.000	0.002	0.179	0.012	0.014
19	Government Share of GDP	GOVSH61	0.007	0.000	0.004	0.177	−0.012	0.016
20	Fraction Buddhist	BUDDHA	0.006	0.000	0.002	0.168	0.009	0.010
27	Higher Education in 1960	H60	0.003	0.000	0.004	0.155	−0.008	0.019
39	Fraction Protestants	PROT00	0.001	0.000	0.000	0.150	−0.005	0.010
41	Population Growth Rate 1960–90	DPOP6090	0.001	0.000	0.010	0.149	0.000	0.015
34	Fraction Population Less than 15	POP1560	0.002	0.000	0.003	0.149	0.001	0.015
36	Defense Spending Share	GDE1	0.002	0.000	0.004	0.148	0.001	0.019
22	Fraction Population Over 65	POP6560	0.005	0.000	0.007	0.148	0.002	0.019
49	Public Educ. Spend. /GDP in 1960s	GEEREC1	0.001	0.000	0.008	0.147	0.004	0.018
58	Gov C Share deflated with GDP prices	GOVNOM1	0.001	0.000	0.001	0.144	−0.006	0.014
43	Fraction Hindus	HINDU00	0.001	0.000	0.001	0.142	0.005	0.012
61	Terms of Trade Growth in 1960s	TOT1DEC1	0.001	0.000	0.001	0.141	0.001	0.015
67	Public Investment Share	GGCFD3	0.000	0.000	0.001	0.139	0.003	0.012
17	Latin American Dummy	LAAM	0.008	0.000	0.001	0.133	−0.005	0.006
50	Fraction Othodox	ORTH00	0.001	0.000	0.001	0.122	0.004	0.010
37	Revolutions and Coups	REVCOUP	0.002	0.000	0.000	0.115	−0.003	0.006
32	Fraction Speaking Foreign Language	OTHFRAC	0.002	0.000	0.000	0.114	0.004	0.005
26	Fraction Muslim	MUSLIM00	0.003	0.000	0.001	0.113	0.000	0.005
42	Civil Liberties	CIV72	0.001	0.000	0.000	0.111	−0.003	0.006
23	Colony Dummy	COLONY	0.004	0.000	0.000	0.108	−0.004	0.004
55	Terms of Trade Ranking	TOTIND	0.001	0.000	0.000	0.105	−0.002	0.007
29	British Colony Dummy	BRIT	0.003	0.000	0.000	0.105	0.004	0.004
56	Fraction Spent in War 1960–90	WARTIME	0.001	0.000	0.000	0.102	−0.002	0.007
35	English Speaking Population	ENGFRAC	0.002	0.000	0.000	0.100	−0.003	0.006
25	European Dummy	EUROPE	0.004	0.000	0.000	0.096	−0.001	0.007
28	Fertility Rates in 1960s	FERTLDC1	0.003	0.000	0.000	0.095	−0.002	0.006
44	Tropical Climate Zone	ZTROPICS	0.001	0.000	0.000	0.094	−0.002	0.006
45	Religion Measure	HERF00	0.001	0.000	0.000	0.093	−0.001	0.006
54	Socialist Dummy	SOCIALIST	0.001	0.000	0.000	0.092	−0.002	0.005
47	Fraction Catholic	CATH00	0.001	0.000	0.000	0.086	−0.002	0.006
65	Oil Producing Country Dummy	OIL	0.000	0.000	0.000	0.084	0.001	0.005
66	Fraction Land Area Near Navig. Water	LT100CR	0.000	0.000	0.000	0.079	−0.002	0.005
53	Outward Orientation	SCOUT	0.001	0.000	0.000	0.075	−0.003	0.003
5	Life Expectancy	LIFE060	0.037	0.000	0.000	0.068	0.001	0.000
63	Landlocked Country Dummy	LANDLOCK	0.001	0.000	0.000	0.065	0.000	0.004
48	War Particpation 1960–90	WARTORN	0.001	0.000	0.000	0.057	−0.001	0.003
62	Timing of Independence	NEWSTATE	0.001	0.000	0.000	0.046	0.001	0.002
33	Political Rights	PRIGHTS	0.002	0.000	0.000	0.042	−0.001	0.001
9	Investment Price	IPRICE1	0.027	0.000	0.000	0.033	0.000	0.000
59	Size of Economy	SIZE60	0.001	0.000	0.000	0.029	0.000	0.001
46	Capitalism	ECORG	0.001	0.000	0.000	0.025	0.000	0.001
51	Hydrocarbon Deposits in 1993	LHCPC	0.001	0.000	0.000	0.009	0.000	0.000
15	Real Exchange Rate Distortions	RERD	0.009	0.000	0.000	0.008	0.000	0.000
16	Absolute Latitude	ABSLATIT	0.008	0.000	0.000	0.008	0.000	0.000
38	Average Inflation 1960–90	PI6090	0.002	0.000	0.000	0.003	0.000	0.000
14	Population Coastal Density	DENS65C	0.009	0.000	0.000	0.001	0.000	0.000
52	Interior Density	DENS65I	0.001	0.000	0.000	0.001	0.000	0.000
57	Population Density	DENS60	0.001	0.000	0.000	0.000	0.000	0.000
31	Square of Inflation 1960–90	SQPI6090	0.002	0.000	0.000	0.000	0.000	0.000
40	Air Distance to Big Cities	AIRDIST	0.001	0.000	0.000	0.000	0.000	0.000
64	Population in 1960	POP60	0.001	0.000	0.000	0.000	0.000	0.000
60	Land Area	LANDAREA	0.001	0.000	0.000	0.000	0.000	0.000

PIP stands for "posterior inclusion probability", PM stands for "posterior mean" and refers to the mean of the posterior distribution of the corresponding slope parameter and PSD stands for "posterior standard deviation" and refers to the standard deviation of the posterior distribution, "LS, 2009" refers to the ordering by PIP using the setting in Ley and Steel (2009). Rows ordered by PIP obtained with the Bayesian elastic net.

Table 1: Estimation results: Bayesian elastic net versus g-prior with spike and slab prior

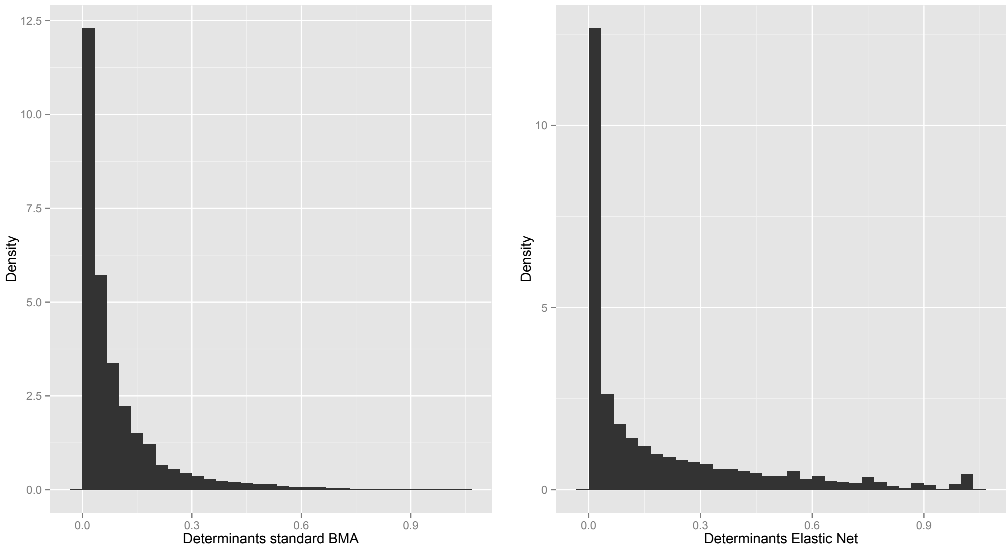


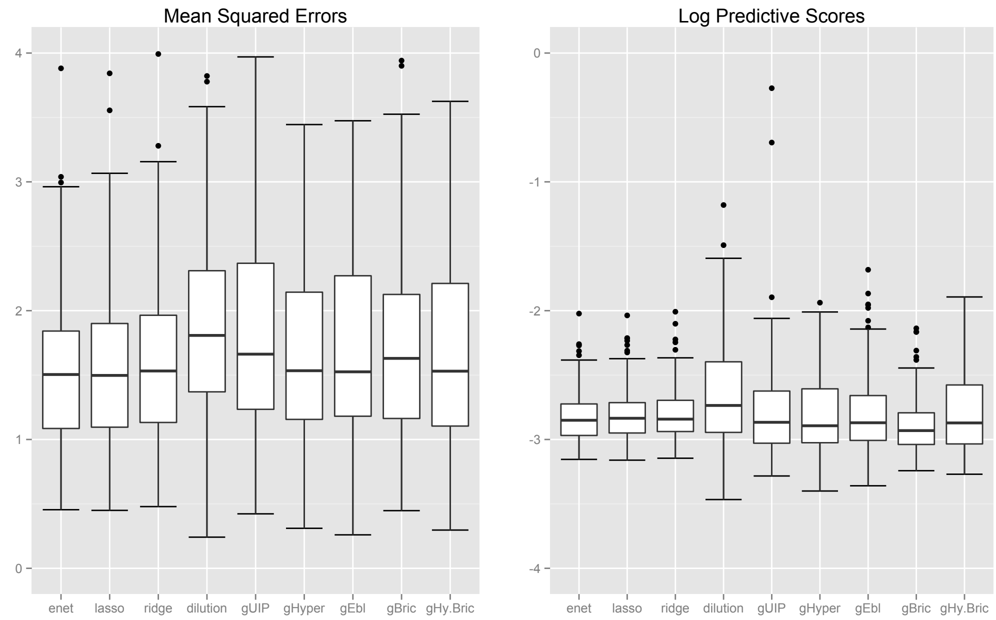
Figure 1: Histogram of the determinants of the regressor correlation matrices of the models visited by the Markov chain in the standard BMA procedure (left) and the Bayesian elastic net (right)

prior settings described in Section 2 making use of an out-of-sample prediction exercise. Using the dataset in Sala-i-Martin et al. (2004), we assign to each observation a probability of 0.15 of belonging to the out-of-sample subsample and, therefore, 0.85 to be part of the in-sample data. For each of the priors, we perform inference based on the observations of the realized in-sample group and obtain point predictions for the out-of-sample subsample. Each point prediction, in turn, is given by the weighted average of the corresponding model-specific conditional expectation, where the weights correspond to the posterior model probabilities obtained using the in-sample observations. We obtain the prediction error for each replication based on the point forecasts from the best 10,000 models in terms of posterior model probability of each model estimated. For each prior setting, we repeat this procedure 100 times.

We evaluate models based on two fixed g-priors (UIP and BRIC—which in our setting is also the RIC prior), the empirical Bayes local prior, two g-hyperpriors (one where the shrinkage prior is chosen to have the expected value corresponding to the UIP prior and another one whose prior expected value corresponds to the BRIC shrinkage), as well as the prior settings leading to the Bayesian LASSO, Bayesian ridge and Bayesian elastic net models. In addition, we also estimated models with the BRIC g-prior and the *dilution prior* proposed by George (1999) (see also George (2010)) to address correlated regressors. George (1999) proposes a modification of the prior over specifications implied by the spike and slab prior in Equation (2). In particular, the method multiplies the standard prior model probability by the determinant of the correlation matrix corresponding to the regressors included in the specification. Such a prior setting introduces therefore an additional penalty for models which include highly correlated covariates (and thus a regressor correlation matrix which is very close to zero). In the spirit of the proposals by George (2010), this specification of the prior over the model space is aimed at overcoming the caveats implied by the independence prior inclusion structure across covariates imposed by standard approaches (among others, the spike and slab prior).

Table 2 and Figure 2 display summary statistics for the prediction errors corresponding to each prior setting. For each replication of the out-of-sample forecasting exercise, we compute the mean squared prediction error (MPE) and the log-predictive score (LPS). The mean squared prediction error is defined as $MPE = \frac{1}{N_F}\sum_{f \in F}(y_f - \mathbf{E}(y_f|\mathbf{y}_{IS}, \mathbf{X}))^2$, where the set F, with N_F observations, defines the observations in the out-of-sample subsample and the in-sample observations are indexed by IS. The log-predictive score, on the other hand, has been proposed as a better indicator of predictive ability in Bayesian model averaging exercises (see for example Fernández et al. (2001a) or Eicher et al. (2011)). The log-predictive score is given by $LPS = -\frac{1}{N_F}\sum_{f \in F}\log p(y_f|y_{IS}, \mathbf{X})$, with lower values indicating better predictive performance.

The characteristics of the mean squared prediction errors differ between the three types of bridge-like regression models (LASSO, ridge and elastic net) and the rest of the specifications. LASSO, elastic net and ridge, in that order, present the best predictive ability as measured by the median of the MPE statistic. The difference in terms of median predictive ability, however, is very small with respect to the rest of the prior specifications. Models with hyperpriors over the g-prior present better forecasting performance than those with fixed g-priors, which in turn outperforms the dilution prior proposed by George (1999). The LPS ranking, on the other hand, presents a better (median) performance of models with g-priors as compared to the group of bridge shrinkage models, with minimal differences across specifications.



Notes: enet: Bayesian elastic net, lasso: Bayesian LASSO, ridge: Bayesian ridge,dilution: risk inflation criterion g-prior and dilution prior over the model space as in George (1999, 2010), gUIP: unit information g-prior, gHyper: g-prior with Beta-hyperprior on the shrinkage factor such that the prior expectation corresponds to the unit information prior, gEbl: g-prior with empirical Bayes local, gBric: risk inflation criterion g-prio, gHy.Bric: g-prior with Beta-hyperprior on the shrinkage factor such that the prior expectation corresponds to the risk inflation criterion priorr.

Figure 2: Box plots of the mean prediction errors (multiplied by 10^4) and log-predictive scores for different priors over the slope parameters, ordered by median

While such small differences do not allow us to infer clear-cut conclusions on predictive abilities based on the median (or average) of the distribution of out-of-sample predictive errors, the dispersion of such errors does lead to a more systematic grouping of alternative prior structures. The interquartile range of prediction errors in the Bayesian LASSO, ridge and elastic net regression is systematically lower independently of whether the MPE or the LPS is used for the comparison of forecasting ability. A comparison of the statistics in Table 2 indicates that the shrinkage implied by bridge-type priors presents better properties in

	Elastic net	LASSO	Ridge	Dilution prior	g-Hyp. UIP	g-EBL	g-Hyp. BRIC	BRIC	UIP
MPE statistics:									
1st Quartile	1.085	1.095	1.132	1.387	1.156	1.187	1.108	1.252	1.271
Median	1.504	1.497	1.532	1.809	1.554	1.548	1.560	1.692	1.714
3rd Quartile	1.841	1.900	1.964	2.405	2.184	2.273	2.266	2.446	2.541
LPS statistics:									
1st Quartile	−2.969	−2.950	−2.938	−2.945	−3.025	−3.007	−3.035	−3.033	−3.026
Median	−2.851	−2.835	−2.842	−2.736	−2.893	−2.870	−2.871	−2.908	−2.853
3rd Quartile	−2.724	−2.714	−2.696	−2.397	−2.607	−2.659	−2.577	−2.713	−2.564

Notes: Elastic net: Bayesian elastic net, LASSO: Bayesian LASSO, Ridge: Bayesian ridge, Dilution prior: risk inflation criterion g-prior and dilution prior over the model space as in George (1999, 2010), g-Hyp. UIP: g-prior with Beta-hyperprior on the shrinkage factor such that the prior expectation corresponds to the unit information prior, g-EBL: g-prior with empirical Bayes local, g-Hyp. BRIC: g-prior with Beta-hyperprior on the shrinkage factor such that the prior expectation corresponds to the risk inflation criterion prior, BRIC: risk inflation criterion g-prior, UIP: unit information g-prior

Table 2: Summary statistics of the mean prediction (squared) errors (MPE, multiplied by 10^4) and log-predictive scores (LPS) by prior setting

terms of the risk embodied in the distribution of predictive errors.

In order to investigate the structure of the predictive loss across observations, we analyze the distribution of the LPS across subsamples defined by world regions.[10] We divide the full sample of countries into *Europe*, *Sub-Saharan Africa*, *East Asia*, *Latin America* (as defined by the dummy variables included in the set of covariates presented in Table 1) and *Rest of the World*, and obtain the distribution of LPS within each one of these subsamples. The corresponding LPS distributions for the subsamples defined above are presented in Figure 3. The comparison of the distribution of LPS across subsamples reveals a large degree of heterogeneity in the predictive accuracy for all the different prior structures. The dispersion of LPS values varies strongly across world region subsamples for all prior elicitation methods, with a particularly large variance for Sub-Saharan Africa and East Asia. The relative ordering of methods in terms of average performance presents a strong variability across samples. While the prior based on dilution leads to sizeable improvements in average out-of-sample predictive performance for European countries, for instance, the LPS distribution for this method in the subsample of Sub-Saharan economies presents much higher first and second moments than those of all other prior settings. The strong heterogeneity in predictive ability across groups countries of methods such as the dilution prior and the Beta-hyperprior contrasts with the relative low and stable variability of the risk inflation criterion, elastic net, ridge and LASSO shrinkage approaches within world regions compared to the rest of the approaches.

The analysis highlights the large variability associated with out-of-sample prediction of economic growth based on cross-sectional subsamples after accounting for model uncertainty. In the framework of standard decision rules based on forecasts, the small differences observed in mean prediction error among prior specifications would imply a preference for methods with relatively low predictive error variance. Our analysis shows that, in this respect, the characteristics of priors which present strong shrinkage (in particular, ridge, LASSO and elastic nets) are particularly suitable for building model-averaged predictions of economic growth based on cross-sectional information.

[10]The results for the MPE are qualitatively similar to those reported for the LPS and therefore not presented here. These are available from the authors upon request.

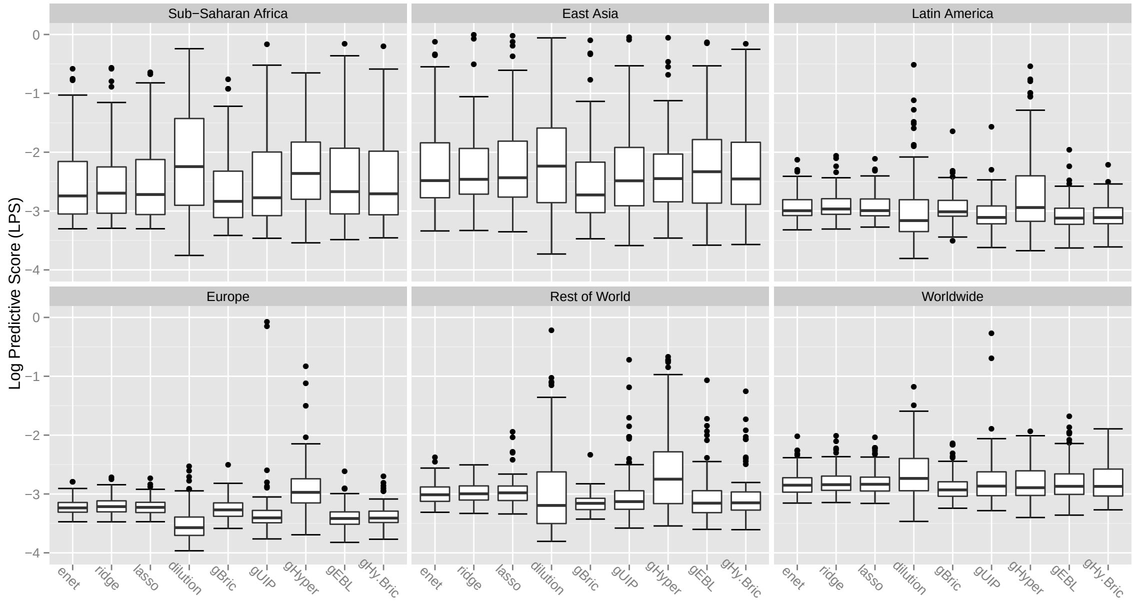


Figure 3: Box plots of the log-predictive scores for different priors over the slope parameters, by world region

With the aim of assessing the sensitivity of the prediction performance to outlying data events, we design an alternative setting based on randomly assigning shocks to some of our explanatory variables. We repeat the out-of-sample prediction exercise presented above, but randomly replacing some of the variables by their realized 25th or 75th percentile. In particular, shocks are assigned to the variables in each one of the replications of the out-of-sample prediction exercise as follows. First, we assign a 25% quantile or a 75% quantile event, each with a 0.5 probability. Such an event is then applied to randomly chosen variables from the group of shocked covariates (with probability 0.5 for each one of them), which is then replaced by the corresponding percentile. We choose the group of variables given by *Malaria Prevalence, Primary School Enrolment, Fraction of Tropical Area, Years Open, Fraction Confucian* and *Life Expectancy* as potentially "shocked" variables.[11] The mean absolute correlation between these variables is 0.543, which contrasts with an absolute correlation of 0.210 for the rest of the covariates considered.

Table 3 illustrates the results of this alternative out-of-sample prediction exercise in terms of MPE and LPS. Bayesian LASSO, elastic net and ridge priors systematically outperform the rest of the methods in terms of prediction accuracy. Specifically, the treatment of correlated variables across priors appears to be the key element to understand the superiority of these shrinkage methods in this altered out-of-sample setting where outlying shocks are implemented. The parameter associated to the *Life Expectancy* variable, for instance, which is the covariate with the highest average correlation to the other shock variables (0.65), is strongly shrunk to zero if Bayesian LASSO, elastic net or ridge estimators are used (see the results in Table 1). On the other hand, alternative shrinking estimators such as the dilution prior identify it as an important covariate with a high degree of jointness with other correlated variables, thus resulting in very high prediction errors under this alternative scenario.

[11]These correspond to the covariates with the highest PIP for the Bayesian elastic net priors (*Malaria Prevalence, Primary School Enrolment, Fraction of Tropical Area, Years Open, Fraction Confucian*), as well as one of the most robust variables in Ley and Steel (2009) (*Life Expectancy*).

	Elastic net	LASSO	Ridge	Dilution prior	g-Hyp. UIP	g-EBL	g-Hyp. BRIC	BRIC	UIP
MPE statistics:									
1st Quartile	3.82	3.90	3.62	31.16	19.06	21.14	19.56	4.61	15.67
Median	5.29	5.32	4.76	43.26	27.34	29.88	28.52	5.87	24.63
3rd Quartile	6.27	6.43	6.47	71.24	47.42	51.36	48.62	7.30	42.15
LPS statistics:									
1stQuartile	−2.93	−2.91	−2.91	−2.50	−2.75	−2.73	−2.77	−2.75	−2.76
Median	−2.62	−2.61	−2.58	−2.12	−2.58	−2.51	−2.58	−2.57	−2.58
3rd Quartile	−2.23	−2.23	−2.22	−1.63	−2.31	−2.28	−2.36	−2.33	−2.34

Table 3: Summary statistics of the mean prediction (squared) errors (MPE, multiplied by 10^4) and the log predictive scores (LPS) in the simulations including outlying shocks

5 Conclusions

Using the cross-country growth regression dataset in Sala-i-Martin et al. (2004), we compare a large set of priors aimed at dealing simultaneously with model uncertainty and correlated regressors in linear regression models. These methods combine spike and slab priors with several approaches to shrinkage, including standard fixed g-priors, specifications with hyperpriors on the g-prior, Bayesian LASSO, Bayesian ridge regression and Bayesian elastic nets. Our results indicate that Bayesian LASSO, Bayesian ridge regression and Bayesian elastic nets present better out-of-sample prediction properties than standard model averaging methods which do not explicitly account for shrinkage in individual specifications beyond the penalty implied by the posterior model probability when Zellner's g-priors are used (Zellner (1986)). Such improvements in out-of-sample prediction materialize in less dispersed distributions of prediction errors and a more predictive ability when covariates are subject to outlying shocks.

Acknowledgments

The authors would like to thank two anonymous referees, Mark Steel as well as participants at the 5th CSDA International Conference on Computational and Financial Econometrics (CFE'11), the Statistische Woche 2012 and the Austrian Economic Association Conference for helpful comments on an earlier version of this paper. An earlier version of this paper concentrating exclusively on Bayesian elastic nets was distributed under the title *"Fishing Economic Growth Determinants Using Bayesian Elastic Nets"*. The first author's research was supported by the Oesterreichische Nationalbank under the Jubiläumsfond grant 14663.

References

Bornn, L., Gottardo, R., and Doucet, A. (2010). Grouping priors and the Bayesian elastic net. Technical Report 254, The University of British Columbia, Department of Statistics.

Brown, P., Vannucci, M., and Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society B*, 60:627–641.

Chipman, H. (1996). Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, 24:17–36.

Crespo Cuaresma, J. (2011). How different is Africa? A comment on Masanjala and Papageorgiou (2008). *Journal of Applied Econometrics*, 26:1041–1047.

Crespo Cuaresma, J. and Doppelhofer, G. (2007). Nonlinearities in cross-country growth regressions: A Bayesian averaging of thresholds (BAT) approach. *Journal of Macroeconomics*, 29:541–554.

Doppelhofer, G. and Weeks, M. (2009). Jointness of growth determinants. *Journal of Applied Econometrics*, 24(2):209–244.

Dunson, D., Herring, A., and Engel, S. (2008). Bayesian selection and clustering of polymorphisms in functionally related genes. *Journal of the American Statistical Association*, 103:534–546.

Durlauf, S., Kourtellos, A., and Tan, C. (2008). Are any growth theories robust? *Economic Journal*, 118:329–346.

Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, 26(1):30–55.

Feldkircher, M. and Zeugner, S. (2009). Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging. IMF Working Paper WP/09/202.

Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100:381–427.

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

Foster, D. P. and George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22:1947–1975.

George, E. (1999). Discussion of Bayesian model averaging and model search strategies by M.A. Clyde. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 6, pages 175–177. Oxford University Press.

George, E. (2010). Dilution priors: Compensating for model space redundancy. *IMS Collections. Borrowing Strength: Theory Powering Applications A Festschrift for Lawrence D. Brown*, 6:158–165.

George, E. and Foster, D. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, 87(4):731–747.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for {VAR} model restrictions. *Journal of Econometrics*, 142(1):553 – 580.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.

Hansen, M. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774.

Hoerl, A. and Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

Jochmann, M., Koop, G., and Strachan, R. W. (2010). Bayesian forecasting using stochastic search variable selection in a {VAR} subject to breaks. *International Journal of Forecasting*, 26(2):326 – 347.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795.

Kass, R. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Korobilis, D. (2013). Bayesian forecasting with highly correlated predictors. *Economics Letters*, 118:148–150.

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5:369–412.

Ley, E. and Steel, M. (2007). Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics*, 29(3):476–493.

Ley, E. and Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions. *Journal of Applied Econometrics*, 24(4):651–674.

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

Li, Q. and Lin, N. (2010). The Bayesian elastic net. *Bayesian Analysis*, 5(1):151–117.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103:410–423.

Mitchell, T. and Beauchamp, J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103:681–686.

Poirier, D. (1995). *Intermediate Statistics and Econometrics—A Comparative Approach*. MIT Press.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279–300.

R Core Team (2012). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

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.

Schneider, U. and Wagner, M. (2012). Catching growth determinants with the adaptive lasso. *German Economic Review*, 13(1):71–85.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288.

Yuan, M. and Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100:1215–1225.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, edited by P. Goel and A. Zellner.