Fishing Economic Growth Determinants Using Bayesian Elastic Nets

Paul Hofmarcher* *JKU*

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

Bettina Grün‡ *JKU*

Kurt Hornik§ *WU*

Abstract

We propose a method to deal simultaneously with model uncertainty and correlated regressors in linear regression models by combining elastic net specifications with a *spike and slab* prior. The estimation method nests ridge regression and the LASSO estimator and thus allows for a more flexible modelling framework than existing model averaging procedures. In particular, the proposed technique has clear advantages when dealing with datasets of (potentially highly) correlated regressors, a pervasive characteristic of the model averaging datasets used hitherto in the econometric literature. We apply our method to the dataset of economic growth determinants by Sala-i-Martin et al. (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) and show that our procedure has superior out-of-sample predictive abilities as compared to the standard Bayesian model averaging methods currently used in the literature.

*Department of Applied Statistics, Johannes Kepler University Linz. Altenbergerstraße 69, 4040 Linz, Austria. paul.hofmarcher@jku.at

†Department of Economics, Vienna University of Economics and Business (WU). Augasse 2-6, 1090 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). Augasse 2-6, 1090 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. (2001), 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) or Eicher et al. (2011) 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 (2007)), a systematic assessment of the issue is hitherto missing.[1] In this paper we propose a method to deal with the problem of model uncertainty in the presence of correlated regressors. The framework of bridge regression allows us to deal explicitly with the problem of correlated explanatory variables by shrinking coefficients. Prominent cases of the bridge regression class are ridge regression and LASSO. Our method is based on Bayesian elastic net specifications, which nest both ridge regression and the LASSO estimator as special cases. Additionally, we propose using a *spike and slab* prior (see for instance Mitchell and Beauchamp (1988) or George and McCulloch (1993)) which allows us to perform variable selection or model averaging in the framework of the Bayesian elastic net. In addition, 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.

We evaluate our method 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 (2008) 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. The use of Bayesian elastic nets leads to some important changes in the results of the robustness analysis as compared to the existing literature. As in Schneider and Wagner (2008), compared to Sala-i-Martin et al. (2004) variables like *Population Coastal Density* or *Life Expectancy* strongly reduce their importance in our results, but variables like *Malaria Prevalence* and *Years Open* appear as more robust growth determinants. In addition, we perform an out-of-sample prediction exercise which confirms the superiority of the Bayesian elastic net with spike and slab priors as compared to

[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).

standard linear Bayesian model averaging methods.

The paper is structured as follows. Section 2 presents the Bayesian elastic net, section 3 explains the approach we take to model uncertainty in the framework of Bayesian elastic net models and section 4 performs the empirical analysis based on Sala-i-Martin et al. (2004). Finally, section 5 concludes.

2 Ridge regression, LASSO and the elastic net

Assume that a group of K variables $\{x_1, \dots x_K\}$ are proposed as potential determinants of y 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 y, $\mathbf{X}$ is the $N \times K$ design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1 \dots \beta_K)'$ denotes the parameter vector of interest and it is assumed that $\mathbf{u} \sim \mathbf{N}(0, \sigma^2 \mathbf{I}_N)$. Assuming that $N > K$, the standard 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 $E((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \sigma^2 \sum_j^K \lambda_j^{-1}$, where $\{\lambda_1, \dots, \lambda_K\}$ are the eigenvalues of $(\mathbf{X}'\mathbf{X})$.[2] 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 regression coefficients are thus obtained as

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

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 (2) 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 (2) 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.

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.

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

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

$$\boldsymbol{\beta}|\boldsymbol{\tau}^2, \sigma^2 \sim \mathbf{N}(0, \sigma^2 \mathbf{W}_{\boldsymbol{\tau}}), \tag{3}$$

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

$$p(\sigma^2) = 1/\sigma^2, \tag{4}$$

and the LASSO estimator is obtained by assigning an independent double exponential (or Laplace) distribution for each τ_j^2. The ridge regression estimator, on the other hand, is obtained by imposing the common inverse gamma distribution as a prior over τ^2.

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

$$\hat{\boldsymbol{\beta}}_{\text{enet}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \sum_{j=1}^{K} \left(\lambda_1 |\beta_j| + \lambda_2 \beta_j^2 \right) \right\}. \tag{5}$$

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

$$\boldsymbol{\beta}|\sigma^2 \sim \exp\left\{ -\frac{1}{2\sigma^2} \left[\lambda_1 \sum_{j=1}^{K} |\beta_j| + \lambda_2 \sum_{j=1}^{K} \beta_j^2 \right] \right\}. \tag{6}$$

This prior over $\boldsymbol{\beta}$ conditional on σ^2, combined with (4) 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. 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.

3 Model uncertainty and the Bayesian elastic net

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 $[-\sqrt{\mathrm{var}(\beta_j|y)}, \sqrt{\mathrm{var}(\beta_j|y)}]$ and a variable is excluded if this posterior probability exceeds a certain ad hoc probability threshold.

In this contribution we further expand the variable selection method by modifying the prior on the β vector. We propose a prior 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 the prior distribution for β_j described above.[3] That implies that the prior on β_j is given by

$$p(\beta_j|\gamma_j, \tau_j, \sigma^2) \sim (1-\gamma_j)I_0 + \gamma_j \pi(\beta_j|\tau_j, \sigma^2) \tag{7}$$

where $\pi(\beta_j|\tau_j, \sigma^2)$ is the prior distribution of β_j implied by (6) after reparametrizing it as a mixture of normals and including the additional parameter vector $\boldsymbol{\tau}$. A Bernoulli prior is assumed on γ_j, so that $\gamma_j \sim \mathrm{Be}(\underline{\gamma})$. 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 Bayesian elastic net 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 by comparing it to 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. (2001), Sala-i-Martin et al. (2004) or Ley and Steel (2009) for empirical applications related to economic growth).

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.

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 $\underline{\gamma} \sim \mathrm{Beta}(a, b)$. 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.

[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)).

We also follow this approach in our empirical application.

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 elastic net. 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 (7).

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)).

4 Empirical application: Fishing economic growth determinants

4.1 Robust economic growth determinants using Bayesian elastic nets

Sala-i-Martin et al. (2004) (henceforth, SDM) study the robustness of economic growth determinants to model uncertainty using a dataset for 88 countries comprising data on GDP 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.[4] 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. 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).

We apply the model to the data using 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 $\lambda^2 \sim \text{Gamma}(0.1, 0.1)$.[5] 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

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

[5] 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.

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 in Sala-i-Martin et al. (2004) with those obtained from estimating the Bayesian elastic net with spike and slab priors on the inclusion of the variables.[6] The first column presents the original ranking in terms of PIP implied by the results of SDM. 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 results presented by SDM and those obtained using the Bayesian elastic net. The PIP of the SDM results and those of the elastic net have a correlation of 0.61, and strong differences can be observed when comparing the relative importance of the variables in the dataset. The mean of the posterior distribution of γ_j 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), who use a comparable hyperprior on the prior inclusion probabilities of the variables, albeit in the framework of standard linear models.

The shrinkage implied by the Bayesian elastic net has a strong effect on the nature of the robust determinants of economic growth implied by the results in Table 1. On the one hand, some of the variables with highest posterior inclusion probability in SDM (in particular *Investment Price*, but also *Population Coastal Density* and *Life Expectancy*) strongly reduce their importance in the results obtained by the Bayesian elastic net. On the other hand, *Malaria Prevalence* and *Years Open* appear as very robust determinants of economic growth in our results and improve their relative importance significantly as compared to the original results in SDM.

The differences in results between the two methods can be traced back to the way that the two model averaging techniques deal with correlated regressors. A standard measure for the degree of collinearity among the variables in a given model is given by the determinant $|\mathbf{R}_k|$ of the correlation matrix of regressors, a measure proposed by George (2007) as a building block of dilution priors over the model space. This determinant equals to one if the columns of $\mathbf{X}_k$ are orthogonal and zero for perfectly collinear columns in $\mathbf{X}_k$. We compute this measure for all models visited by the Markov chain for each one of the two methods evaluated and the histograms of $|\mathbf{R}_k|$ 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.

These results indicate that in this application the Bayesian elastic net leads 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.

[6] All the computations within this work are done by using R, a language and environment for statistical computing (R Development Core Team (2011)) and its extension packages rjags, coda and bms. Codes are available from the authors upon request.

# (BACE)	Description	Name	BACE PIP	BACE PM	BACE PSD	Elastic net PIP	Elastic net PM	Elastic net PSD
1	East Asian Dummy	EAST	0.8225	0.0218	0.0061	0.9756	0.0239	0.0065
7	Malaria Prevalence	MALFAL66	0.2519	-0.0158	0.0062	0.6566	-0.0151	0.0064
2	Primary Schooling Enrollment	P60	0.7965	0.0268	0.0080	0.5902	0.0171	0.0078
5	Fraction of Tropical Area	TROPICAR	0.5633	-0.0148	0.0042	0.3672	-0.0101	0.0052
14	Years Open 1950-94	YRSOPEN	0.1193	0.0124	0.0063	0.3638	0.0118	0.0063
9	Fraction Confucian	CONFUC	0.2065	0.0538	0.0220	0.2291	0.0182	0.0190
13	Spanish Colony Dummy	SPAIN	0.1232	-0.0109	0.0050	0.2154	-0.0078	0.0052
18	Government Consumption Share	GVR61	0.1038	-0.0452	0.0239	0.2085	-0.0165	0.0190
4	Initial Income (Log GDP in 1960)	GDPCH60L	0.6846	-0.0085	0.0029	0.2048	-0.0059	0.0034
17	Ethnolinguistic Fractionalization	AVELF	0.1047	-0.0114	0.0059	0.2023	-0.0088	0.0065
22	(Imports + Exports)/GDP	OPENDEC1	0.0761	0.0089	0.0053	0.1996	0.0072	0.0049
27	Primary Exports in 1970	PRIEXP70	0.0533	-0.0115	0.0076	0.1914	-0.0086	0.0068
26	Fraction Population In Tropics	TROPPOP	0.0580	-0.0108	0.0069	0.1895	-0.0087	0.0068
10	Sub-Saharan Africa Dummy	SAFRICA	0.1537	-0.0153	0.0071	0.1847	-0.0080	0.0062
12	Fraction GDP in Mining	MINING	0.1241	0.0375	0.0184	0.1791	0.0119	0.0143
24	Government Share of GDP	GOVSH61	0.0632	-0.0373	0.0263	0.1768	-0.0119	0.0162
16	Fraction Buddhist	BUDDHA	0.1084	0.0212	0.0109	0.1682	0.0093	0.0097
25	Higher Education in 1960	H60	0.0614	-0.0709	0.0412	0.1555	-0.0084	0.0189
29	Fraction Protestants	PROT00	0.0458	-0.0123	0.0097	0.1500	-0.0050	0.0101
56	Population Growth Rate 1960-90	DPOP6090	0.0192	0.0471	0.3015	0.1494	-0.0003	0.0152
31	Fraction Population Less than 15	POP1560	0.0411	0.0447	0.0406	0.1487	0.0012	0.0147
45	Defense Spending Share	GDE1	0.0213	0.0523	0.0700	0.1482	0.0011	0.0193
44	Fraction Population Over 65	POP6560	0.0224	0.0209	0.1179	0.1480	0.0015	0.0193
48	Public Educ. Spend. /GDP in 1960s	GEEREC1	0.0208	0.1281	0.1697	0.1474	0.0038	0.0182
33	Gov C Share deflated with GDP prices	GOVNOM1	0.0355	-0.0335	0.0273	0.1438	-0.0061	0.0140
30	Fraction Hindus	HINDU00	0.0450	0.0188	0.0127	0.1422	0.0047	0.0122
47	Terms of Trade Growth in 1960s	TOT1DEC1	0.0212	0.0349	0.0465	0.1407	0.0008	0.0150
28	Public Investment Share	GGCFD3	0.0482	-0.0578	0.0427	0.1393	0.0029	0.0124
11	Latin American Dummy	LAAM	0.1493	-0.0132	0.0058	0.1333	-0.0050	0.0055
65	Fraction Othodox	ORTH00	0.0151	0.0053	0.0136	0.1221	0.0038	0.0097
41	Revolutions and Coups	REVCOUP	0.0286	-0.0068	0.0062	0.1151	-0.0025	0.0055
21	Fraction Speaking Foreign Language	OTHFRAC	0.0799	0.0069	0.0040	0.1139	0.0045	0.0051
15	Fraction Muslim	MUSLIM00	0.1145	0.0125	0.0061	0.1133	0.0000	0.0054
40	Civil Liberties	CIV72	0.0288	-0.0075	0.0072	0.1113	-0.0025	0.0058
39	Colony Dummy	COLONY	0.0293	-0.0049	0.0047	0.1079	-0.0042	0.0044
63	Terms of Trade Ranking	TOTIND	0.0156	-0.0038	0.0097	0.1053	-0.0024	0.0074
42	British Colony Dummy	BRIT	0.0267	0.0038	0.0036	0.1049	0.0040	0.0035
60	Fraction Spent in War 1960-90	WARTIME	0.0159	-0.0017	0.0092	0.1020	-0.0017	0.0070
53	English Speaking Population	ENGFRAC	0.0200	-0.0028	0.0067	0.0995	-0.0034	0.0055
37	European Dummy	EUROPE	0.0300	-0.0014	0.0100	0.0959	-0.0006	0.0065
36	Fertility Rates in 1960s	FERTLDC1	0.0308	-0.0068	0.0100	0.0945	-0.0018	0.0062
62	Tropical Climate Zone	ZTROPICS	0.0157	-0.0029	0.0066	0.0937	-0.0022	0.0059
50	Religion Measure	HERF00	0.0205	-0.0050	0.0075	0.0931	-0.0010	0.0056
52	Socialist Dummy	SOCIALIST	0.0203	0.0040	0.0050	0.0918	-0.0019	0.0052
35	Fraction Catholic	CATH00	0.0331	-0.0077	0.0093	0.0862	-0.0022	0.0055
55	Oil Producing Country Dummy	OIL	0.0194	0.0041	0.0065	0.0837	0.0014	0.0053
58	Fraction Land Area Near Navig. Water	LT100CR	0.0187	-0.0028	0.0059	0.0787	-0.0015	0.0048
38	Outward Orientation	SCOUT	0.0297	-0.0034	0.0027	0.0755	-0.0025	0.0029
8	Life Expectancy	LIFE060	0.2086	0.0008	0.0004	0.0675	0.0007	0.0003
49	Landlocked Country Dummy	LANDLOCK	0.0205	-0.0023	0.0042	0.0652	0.0003	0.0038
66	War Particpation 1960-90	WARTORN	0.0150	-0.0008	0.0030	0.0565	-0.0011	0.0029
57	Timing of Independence	NEWSTATE	0.0191	0.0010	0.0020	0.0459	0.0011	0.0020
23	Political Rights	PRIGHTS	0.0656	-0.0018	0.0013	0.0415	-0.0013	0.0014
3	Investment Price	IPRICE1	0.7735	-0.0001	0.0000	0.0328	-0.0001	0.0000
51	Size of Economy	SIZE60	0.0203	-0.0005	0.0014	0.0290	-0.0003	0.0013
64	Capitalism	ECORG	0.0151	-0.0003	0.0011	0.0248	0.0004	0.0011
43	Hydrocarbon Deposits in 1993	LHCPC	0.0246	0.0003	0.0004	0.0089	0.0000	0.0003
20	Real Exchange Rate Distortions	RERD	0.0815	-0.0001	0.0000	0.0084	-0.0001	0.0001
34	Absolute Latitude	ABSLATIT	0.0331	0.0001	0.0002	0.0080	0.0002	0.0002
54	Average Inflation 1960-90	PI6090	0.0197	-0.0001	0.0001	0.0031	-0.0001	0.0001
6	Population Coastal Density	DENS65C	0.4284	0.0000	0.0000	0.0006	0.0000	0.0000
67	Interior Density	DENS65I	0.0150	0.0000	0.0000	0.0006	0.0000	0.0000
19	Population Density	DENS60	0.0861	0.0000	0.0000	0.0004	0.0000	0.0000
59	Square of Inflation 1960-90	SQPI6090	0.0177	0.0000	0.0000	0.0000	0.0000	0.0000
32	Air Distance to Big Cities	AIRDIST	0.0394	0.0000	0.0000	0.0000	0.0000	0.0000
46	Population in 1960	POP60	0.0212	0.0000	0.0000	0.0000	0.0000	0.0000
61	Land Area	LANDAREA	0.0158	0.0000	0.0000	0.0000	0.0000	0.0000

PIP stands for "posterior inclusion probability", PM stands for "posterior mean" and PSD stands for "posterior standard deviation", "# (BACE)" refers to the ordering by PIP in Sala-i-Martin et al. (2004). Rows ordered by PIP obtained from the elastic net.

Table 1: Estimation results: Elastic net versus BACE

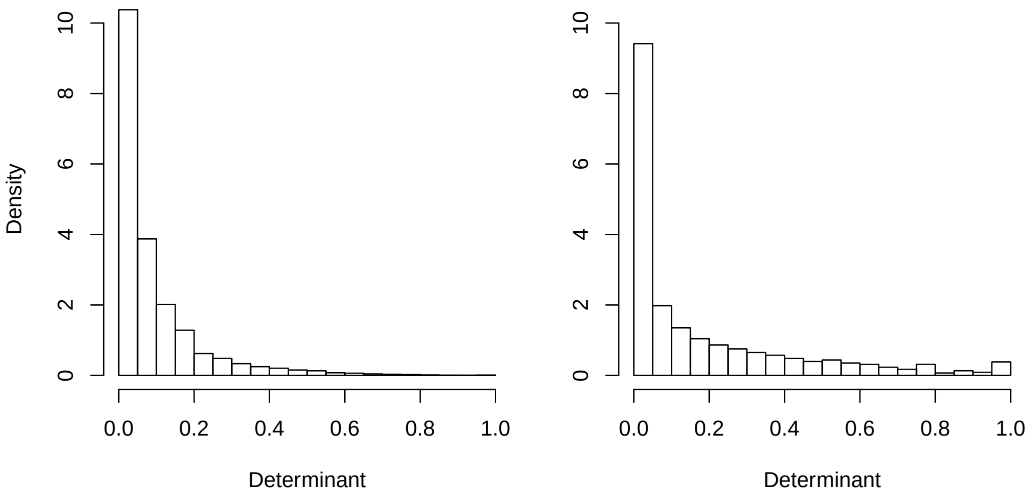

Figure 1: Histogram on the determinant of the regressor correlation matrix for models visited by the Markov chain in the standard BMA procedure (left) and the Bayesian elastic net (right).

4.2 Assessing out-of-sample predictive ability

The different estimation method and shrinkage philosophy can explain the differences in results between standard linear approaches and the Bayesian elastic net presented in Table 1. The superiority of one of the two approaches, however, needs to be assessed in terms of predictive accuracy. For this purpose, we perform an out-of-sample prediction simulation based on the SDM data. We assign to each observation a probability of 0.15 to belong to the out-of-sample group and, therefore, 0.85 to be part of the in-sample data. We then perform inference based on the observations of the realized in-sample group and obtain point predictions for the out-of-sample observations, which are in turn 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 repeat this procedure 100 times, estimating in each replication the Bayesian elastic net and the standard linear counterpart. For the linear model we adopt a fully Bayesian approach, thus deviating slightly from SDM and instead using the approach put forward by Fernández et al. (2001) and expanded by Ley and Steel (2009). We obtain the mean prediction error for each replication based on the best 10,000 models in terms of posterior model probability.

Table 4 and figure 2 display summary statistics of the resulting mean squared prediction errors. Both the average of the out-of-sample prediction error and its dispersion are smaller in the case of the Bayesian elastic net as compared to the standard linear Bayesian model averaging method.

4.3 Robustness checks: LASSO and ridge specifications

We conducted some robustness checks in the framework of the Bayesian elastic net changing priors over the parameters in the model, which left the results presented in

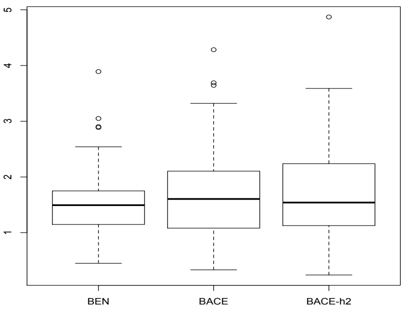

Figure 2: Boxplot of the mean prediction errors (multiplied by 10^4), Bayesian elastic net (BEN) versus standard Bayesian model averaging (BACE) and BACE with hyperpriors on the model size and the coefficients (BACE-h2).

	BEN	BACE	BACE-h2
Min.	0.4516	0.3336	0.2414
1st Qu.	1.1733	1.0865	1.1365
Median	1.4942	1.6053	1.5410
Mean	1.5914	1.6697	1.7220
3rd Qu.	1.7452	2.0981	2.2376
Max.	3.8911	4.2835	4.8706

Table 2: Summary statistics of the mean prediction errors (multiplied by 10^4). Bayesian elastic net (BEN) versus standard Bayesian model averaging (BACE) and BACE-h2 .

Table 1 qualitatively unchanged.[7] We also estimated the models using exclusively Bayesian LASSO and ridge specifications, corresponding to the Bayesian elastic net model presented above with $\lambda_1 = 0$ or $\lambda_2 = 0$ in (6) for, respectively, the LASSO and ridge regression. The results from the estimation for the SDM dataset are presented in Tables 3 and 4. The results for LASSO and ridge regressions are qualitatively very similar to those presented for the Bayesian elastic net, but a couple of interesting differences should be pointed out. While the top variables in terms of PIP are left unchanged across estimation settings, the PIP assigned to *Initial Income (Log GDP in 1960)* in the ridge regression specification is much lower than using other estimation methods. This indicates that, although the overall results concerning the most robust growth determinants are left unchanged when different shrinkage methods are used, the type of shrinkage may have sizeable effects on the relative importance of correlated covariates.

5 Conclusions

We propose a method to deal simultaneously with model uncertainty and correlated regressors in linear regression model and apply it to the cross-country growth regression dataset in Sala-i-Martin et al. (2004). The method is a straightforward generalization of Bayesian elastic nets using spike and slab priors to account for beliefs concerning model size and the relative *a priori* importance of different potential determinants. Our specification presents better out-of-sample prediction abilities 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)).

The method proposed is simple to estimate and presents a high degree of flexibility when setting prior structures. Our results indicate that explicitly assessing the

[7]Detailed results of the robustness checks are available from the authors upon request.

correlation across covariates using shrinkage methods can lead to improvements in modelling economic processes which are subject to model uncertainty. Further assessment of shrinkage methods and priors over the model space could be particularly relevant in the setting of Bayesian elastic nets.

Acknowledgments

The authors would like to thank Mark Steel and participants at the 4th International Conference of the ERCIM Working Group on Computing & Statistics (ERCIM'11) for helpful comments on an earlier version of this paper. The first author's research is 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*, forthcoming.

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.

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.

Fernández, C., Ley, E., and Steel, M. F. (2001). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16:563–576.

George, E. (2007). Discussion of Bayesian Model Averaging and Model Search Strategies by M.A. Clyde. *Bayesian Statistics*, pages 157–77.

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.

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

Hoerl, A. and Kennard, R. (1970). Ridge regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12:55–67.

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.

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

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 Development Core Team (2011). *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. (2008). Catching Growth Determinants with the Adaptive LASSO. *Economic Series, Institute for Advanced Studies.*, 232.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Zellner, A. (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam.

# (BACE)	# (BEN)	Description	Name	PIP	PM	PSD
1	1	East Asian Dummy	EAST	0.9700	0.0247	0.0065
7	2	Malaria Prevalence	MALFAL66	0.6601	-0.0158	0.0065
2	3	Primary Schooling Enrollment	P60	0.5568	0.0173	0.0080
5	4	Fraction of Tropical Area	TROPICAR	0.3483	-0.0104	0.0053
14	5	Years Open 1950-94	YRSOPEN	0.3198	0.0119	0.0067
9	6	Fraction Confucian	CONFUC	0.2205	0.0224	0.0226
13	7	Spanish Colony Dummy	SPAIN	0.2018	-0.0169	0.0175
18	8	Government Consumption Share	GVR61	0.1969	-0.0109	0.0094
17	10	Ethnolinguistic Fractionalization	AVELF	0.1825	-0.0088	0.0066
26	13	Fraction Population In Tropics	TROPPOP	0.1771	-0.0008	0.0059
4	9	Initial Income (Log GDP in 1960)	GDPCH60L	0.1747	-0.0033	0.0047
22	11	(Imports + Exports)/GDP	OPENDEC1	0.1740	-0.0048	0.0064
27	12	Primary Exports in 1970	PRIEXP70	0.1731	-0.0073	0.0051
10	14	Sub-Saharan Africa Dummy	SAFRICA	0.1655	-0.0110	0.0123
12	16	Fraction GDP in Mining	MINING	0.1640	0.0009	0.0137
24	15	Government Share of GDP	GOVSH61	0.1616	0.0009	0.0137
16	17	Fraction Buddhist	BUDDHA	0.1516	0.0100	0.0104
25	18	Higher Education in 1960	H60	0.1459	-0.0110	0.0228
56	20	Population Growth Rate 1960-90	DPOP6090	0.1392	0.0014	0.0318
48	24	Public Educ. Spend. /GDP in 1960s	GEEREC1	0.1368	0.0014	0.0311
31	21	Fraction Population Less than 15	POP1560	0.1359	0.0042	0.0229
44	23	Fraction Population Over 65	POP6560	0.1353	0.0021	0.0255
45	22	Defense Spending Share	GDE1	0.1335	0.0064	0.0206
29	19	Fraction Protestants	PROT00	0.1307	-0.0043	0.0138
33	25	Gov C Share deflated with GDP prices	GOVNOM1	0.1293	-0.0067	0.0108
30	26	Fraction Hindus	HINDU00	0.1276	0.0028	0.0111
47	27	Terms of Trade Growth in 1960s	TOT1DEC1	0.1256	0.0052	0.0138
28	28	Public Investment Share	GGCFD3	0.1241	-0.0015	0.0160
11	29	Latin American Dummy	LAAM	0.1194	-0.0052	0.0055
65	30	Fraction Othodox	ORTH00	0.1095	0.0042	0.0098
41	31	Revolutions and Coups	REVCOUP	0.1031	-0.0050	0.0060
15	33	Fraction Muslim	MUSLIM00	0.1017	0.0048	0.0057
21	32	Fraction Speaking Foreign Language	OTHFRAC	0.1002	0.0045	0.0046
40	34	Civil Liberties	CIV72	0.0967	-0.0046	0.0050
39	35	Colony Dummy	COLONY	0.0955	-0.0041	0.0058
63	36	Terms of Trade Ranking	TOTIND	0.0939	-0.0008	0.0065
42	37	British Colony Dummy	BRIT	0.0929	0.0020	0.0038
60	38	Fraction Spent in War 1960-90	WARTIME	0.0893	-0.0019	0.0074
53	39	English Speaking Population	ENGFRAC	0.0862	-0.0034	0.0056
37	40	European Dummy	EUROPE	0.0840	-0.0007	0.0065
50	43	Religion Measure	HERF00	0.0830	-0.0013	0.0062
52	44	Socialist Dummy	SOCIALIST	0.0821	0.0003	0.0053
36	41	Fertility Rates in 1960s	FERTLDC1	0.0814	-0.0009	0.0058
62	42	Tropical Climate Zone	ZTROPICS	0.0804	-0.0007	0.0058
35	45	Fraction Catholic	CATH00	0.0749	-0.0013	0.0055
55	46	Oil Producing Country Dummy	OIL	0.0719	0.0004	0.0053
58	47	Fraction Land Area Near Navig. Water	LT100CR	0.0684	-0.0010	0.0037
38	48	Outward Orientation	SCOUT	0.0668	-0.0024	0.0034
8	49	Life Expectancy	LIFE060	0.0626	-0.0001	0.0010
49	50	Landlocked Country Dummy	LANDLOCK	0.0581	0.0004	0.0039
66	51	War Particpation 1960-90	WARTORN	0.0494	-0.0012	0.0029
57	52	Timing of Independence	NEWSTATE	0.0394	0.0011	0.0021
23	53	Political Rights	PRIGHTS	0.0354	-0.0014	0.0014
3	54	Investment Price	IPRICE1	0.0276	-0.0001	0.0000
51	55	Size of Economy	SIZE60	0.0240	-0.0003	0.0013
64	56	Capitalism	ECORG	0.0208	0.0004	0.0012
34	59	Absolute Latitude	ABSLATIT	0.0091	0.0001	0.0002
20	58	Real Exchange Rate Distortions	RERD	0.0083	0.0000	0.0002
43	57	Hydrocarbon Deposits in 1993	LHCPC	0.0075	0.0000	0.0003
54	60	Average Inflation 1960-90	PI6090	0.0029	-0.0001	0.0001
6	61	Population Coastal Density	DENS65C	0.0012	0.0000	0.0000
67	62	Interior Density	DENS65I	0.0004	0.0000	0.0000
19	63	Population Density	DENS60	0.0003	0.0000	0.0000
59	64	Square of Inflation 1960-90	SQPI6090	0.0001	0.0000	0.0000
32	65	Air Distance to Big Cities	AIRDIST	0.0000	0.0000	0.0000
46	66	Population in 1960	POP60	0.0000	0.0000	0.0000
61	67	Land Area	LANDAREA	0.0000	0.0000	0.0000

PIP stands for "posterior inclusion probability", PM stands for "posterior mean" and PSD stands for "posterior standard deviation", "# (BACE)" refers to the ordering by PIP in Sala-i-Martin et al. (2004), "# (BEN)" refers to the ordering by PIP according to the Bayesian elastic net. Rows ordered by PIP obtained from the LASSO with a spike and slab prior.

Table 3: Estimation results: LASSO with a spike and slab prior

# (BACE)	# (BEN)	Description	Name	PIP	PM	PSD
1	1	East Asian Dummy	EAST	0.9584	0.0268	0.0061
7	2	Malaria Prevalence	MALFAL66	0.6610	-0.0188	0.0057
2	3	Primary Schooling Enrollment	P60	0.4620	0.0195	0.0075
5	4	Fraction of Tropical Area	TROPICAR	0.2701	-0.0128	0.0049
14	5	Years Open 1950-94	YRSOPEN	0.2052	0.0144	0.0067
9	6	Fraction Confucian	CONFUC	0.1906	-0.0159	0.0227
18	10	Government Consumption Share	GVR61	0.1878	0.0162	0.0226
13	8	Spanish Colony Dummy	SPAIN	0.1409	-0.0102	0.0049
24	15	Government Share of GDP	GOVSH61	0.1273	-0.0235	0.0215
12	16	Fraction GDP in Mining	MINING	0.1190	0.0214	0.0170
27	13	Primary Exports in 1970	PRIEXP70	0.1094	-0.0148	0.0124
25	18	Higher Education in 1960	H60	0.1085	-0.0196	0.0228
26	12	Fraction Population In Tropics	TROPPOP	0.1018	-0.0079	0.0147
17	9	Ethnolinguistic Fractionalization	AVELF	0.0993	-0.0120	0.0071
44	24	Fraction Population Over 65	POP6560	0.0981	0.0064	0.0376
16	17	Fraction Buddhist	BUDDHA	0.0968	0.0009	0.0248
56	20	Population Growth Rate 1960-90	DPOP6090	0.0960	0.0021	0.0391
48	21	Public Educ. Spend. /GDP in 1960s	GEEREC1	0.0949	0.0105	0.0293
45	23	Defense Spending Share	GDE1	0.0925	0.0132	0.0222
31	22	Fraction Population Less than 15	POP1560	0.0913	0.0124	0.0263
22	11	(Imports + Exports)/GDP	OPENDEC1	0.0879	0.0092	0.0051
10	14	Sub-Saharan Africa Dummy	SAFRICA	0.0818	-0.0122	0.0173
47	27	Terms of Trade Growth in 1960s	TOT1DEC1	0.0798	0.0064	0.0253
33	25	Gov C Share deflated with GDP prices	GOVNOM1	0.0788	-0.0074	0.0191
28	28	Public Investment Share	GGCFD3	0.0778	-0.0012	0.0174
30	26	Fraction Hindus	HINDU00	0.0762	0.0072	0.0112
4	7	Initial Income (Log GDP in 1960)	GDPCH60L	0.0685	-0.0062	0.0033
11	29	Latin American Dummy	LAAM	0.0635	-0.0075	0.0063
29	19	Fraction Protestants	PROT00	0.0615	-0.0098	0.0074
65	30	Fraction Othodox	ORTH00	0.0578	0.0087	0.0132
15	34	Fraction Muslim	MUSLIM00	0.0512	0.0083	0.0074
41	32	Revolutions and Coups	REVCOUP	0.0482	-0.0078	0.0068
60	38	Fraction Spent in War 1960-90	WARTIME	0.0399	-0.0040	0.0100
63	35	Terms of Trade Ranking	TOTIND	0.0376	-0.0030	0.0098
40	33	Civil Liberties	CIV72	0.0369	-0.0047	0.0070
39	36	Colony Dummy	COLONY	0.0361	-0.0044	0.0062
37	40	European Dummy	EUROPE	0.0357	-0.0043	0.0062
21	31	Fraction Speaking Foreign Language	OTHFRAC	0.0349	0.0039	0.0057
36	41	Fertility Rates in 1960s	FERTLDC1	0.0343	-0.0009	0.0070
53	39	English Speaking Population	ENGFRAC	0.0336	-0.0048	0.0070
42	37	British Colony Dummy	BRIT	0.0317	0.0027	0.0046
50	42	Religion Measure	HERF00	0.0309	-0.0013	0.0067
62	44	Tropical Climate Zone	ZTROPICS	0.0301	-0.0024	0.0078
52	43	Socialist Dummy	SOCIALIST	0.0295	0.0024	0.0063
35	45	Fraction Catholic	CATH00	0.0275	-0.0017	0.0074
55	46	Oil Producing Country Dummy	OIL	0.0251	0.0006	0.0065
58	47	Fraction Land Area Near Navig. Water	LT100CR	0.0233	-0.0028	0.0062
38	48	Outward Orientation	SCOUT	0.0222	-0.0035	0.0031
49	50	Landlocked Country Dummy	LANDLOCK	0.0179	0.0014	0.0046
66	51	War Particpation 1960-90	WARTORN	0.0146	-0.0016	0.0033
57	52	Timing of Independence	NEWSTATE	0.0105	0.0010	0.0013
23	53	Political Rights	PRIGHTS	0.0093	-0.0009	0.0018
8	49	Life Expectancy	LIFE060	0.0086	0.0002	0.0012
51	55	Size of Economy	SIZE60	0.0061	-0.0007	0.0013
64	56	Capitalism	ECORG	0.0047	0.0003	0.0012
3	54	Investment Price	IPRICE1	0.0043	-0.0001	0.0000
34	59	Absolute Latitude	ABSLATIT	0.0020	0.0002	0.0002
43	57	Hydrocarbon Deposits in 1993	LHCPC	0.0016	-0.0001	0.0004
20	58	Real Exchange Rate Distortions	RERD	0.0011	-0.0001	0.0000
54	60	Average Inflation 1960-90	PI6090	0.0006	-0.0001	0.0001
6	61	Population Coastal Density	DENS65C	0.0001	0.0000	0.0000
67	62	Interior Density	DENS65I	0.0000	0.0000	0.0000
19	63	Population Density	DENS60	0.0000	0.0000	0.0000
32	65	Air Distance to Big Cities	AIRDIST	0.0000	0.0000	0.0000
61	67	Land Area	LANDAREA	0.0000	0.0000	0.0000
46	66	Population in 1960	POP60	0.0000	0.0000	0.0000
59	64	Square of Inflation 1960-90	SQPI6090	0.0000	0.0000	0.0000

PIP stands for "posterior inclusion probability", PM stands for "posterior mean" and PSD stands for "posterior standard deviation", "# (BACE)" refers to the ordering by PIP in Sala-i-Martin et al. (2004), "# (BEN)" refers to the ordering by PIP according to the Bayesian elastic net. Rows ordered by PIP obtained from the Bayesian ridge regression model with a spike and slab prior.

Table 4: Estimation results: Bayesian ridge regression with a spike and slab prior