

**CentER** 

Discussion Paper

No. 2011-082

**BAYESIAN MODEL AVERAGING AND WEIGHTED AVERAGE  
LEAST SQUARES: EQUIVARIANCE, STABILITY, AND  
NUMERICAL ISSUES**

**By Giuseppe De Luca, Jan R. Magnus**

July 21, 2011

ISSN 0924-7815

# Bayesian model averaging and weighted average least squares: equivariance, stability, and numerical issues\*

Giuseppe De Luca  
ISFOL

Jan R. Magnus  
Tilburg University

This version: July 21, 2011

## Abstract

This article is concerned with the estimation of linear regression models with uncertainty about the choice of the explanatory variables. We introduce the Stata commands `bma` and `wals` which implement, respectively, the exact Bayesian Model Averaging (BMA) estimator and the Weighted Average Least Squares (WALS) estimator developed by Magnus et al. (2010). Unlike standard pretest estimators which are based on some preliminary diagnostic test, these model averaging estimators provide a coherent way of making inference on the regression parameters of interest by taking into account the uncertainty due to both the estimation and the model selection steps. Special emphasis is given to a number practical issues that users are likely to face in applied work: equivariance to certain transformations of the explanatory variables, stability, accuracy, computing speed and out-of-memory problems. Performances of our `bma` and `wals` commands are illustrated using simulated data and empirical applications from the literature on model averaging estimation.

*Keywords:* Model uncertainty, Model averaging, Bayesian analysis, Exact computation.

*JEL codes:* C11, C51, C52

---

\* We thank Valentino Dardanoni, Salvatore Modica and Franco Peracchi for extensive and insightful discussions. The SHARE data collection has been primarily funded by the European Commission through the 5th, 6th and 7th framework programmes. Additional funding from the U.S. National Institute on Aging as well as from various national sources is gratefully acknowledged (see [www.share-project.org](http://www.share-project.org) for a full list of funding institutions).

# 1 Introduction

Economic theory provides, in general, some information on the empirical model specification but it offers little guidance on how to specify the exact data generating process for the outcome of interest. The lack of a one-to-one link between theory and empirical model specification thus generates uncertainty regarding, for example, which explanatory variables must be included in the model, which functional forms are appropriate, or which lag length captures dynamic responses. In econometrics, all these problems are known as problems of model uncertainty. Standard econometric practice consists of using the same data for model selection and for estimation and ignoring that the resulting estimators are in fact pretest estimators (i.e. estimators based on some preliminary diagnostic test) and hence traditional statistical theory is not directly applicable. As shown by Magnus and Durbin (1999) and Danilov and Magnus (2004), the model selection process matters and it is likely to have non-negligible effects on the statistical properties of our estimators.

This article is concerned with model uncertainty in the context of linear regression models. We focus on uncertainty about the choice of the explanatory variables because this representation of the problem is also suitable for many other forms of model uncertainty. Following Danilov and Magnus (2004), we distinguish between focus regressors that are always included in the model and auxiliary regressors of which we are less certain. Model uncertainty arises because different subsets of auxiliary regressors can be excluded from the model to improve the statistical properties of the estimated focus parameters (for example their mean squared error). One of the main attractions of the Bayesian model averaging techniques discussed in this article is that they provide a coherent method of inference on the regression parameters of interest by taking explicit account of the uncertainty due to both the estimation and the model selection steps. The literature on model averaging estimation is vast and we refer the reader to Hoeting et al. (1999) for a general discussion. Here, our attention is focused on the exact Bayesian Model Averaging (BMA) estimator developed by Leamer (1978, Chapter 4, Sections 4–6) and the Weighted Average Least Squares (WALS) estimator developed by Magnus et al. (2010). The basic idea of these estimators is computing a weighted average of the conditional estimates across all possible models because each of them provides some information on the focus regression parameters. In the spirit of Bayesian inference, the weight given to each model and the conditional estimates of its parameters are determined on the basis of data and priors.

Although our Stata implementation of BMA and WALS is based on the original Matlab commands associated with Magnus et al. (2010), the new Stata commands `bma` and `wals` also introduce

some improvements. Specifically, our `bma` command is more stable, is considerably faster and requires much less memory than the original Matlab command. As for WALS, we modified the estimation procedure by introducing a preliminary scaling of the explanatory variables. The aim of this preliminary scaling step is twofold: making the WALS estimator scale-equivariant and improving accuracy of the WALS estimates. In addition, our `wals` command is more flexible than the original Matlab command in the specification of the prior distributions.<sup>1</sup>

The remainder of this article is organized as follows. Section 2 formalizes our statistical framework. Section 3 describes the theoretical background of BMA and WALS estimators. Section 4 discusses the property of equivariance with respect to ordering, centering and scale transformations of the explanatory variables. Section 5 describes the syntax of our `bma` and `wals` commands, while Section 6 provides some additional remarks on the Stata and Matlab commands for BMA and WALS. Sections 7, 8 and 9 illustrate the performances of our `bma` and `wals` commands using simulated data and empirical applications from the literature on model averaging estimation. Finally, Section 10 offers some conclusions.

## 2 The statistical framework

Our statistical framework is a linear regression model of the form

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}, \quad (1)$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of observations on the outcome of interest, the  $\mathbf{X}_j$ ,  $j = 1, 2$ , are  $n \times k_j$  matrices of observations on two subsets of deterministic regressors, the  $\boldsymbol{\beta}_j$  are  $k_j \times 1$  vectors of unknown regression parameters, and  $\mathbf{u}$  is an  $n \times 1$  random vector of unobservable disturbances whose elements are i.i.d.  $N(0, \sigma^2)$ . We assume that  $k_1 \geq 1$ ,  $k_2 \geq 0$ ,  $k = k_1 + k_2 \leq n - 1$  and that the design matrix  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$  has full column-rank  $k$ . The reason for partitioning the design matrix  $\mathbf{X}$  in two subsets of regressors is that  $\mathbf{X}_1$  contains explanatory variables which we want in the model because of theoretical reasons or other considerations about the phenomenon under investigation, while  $\mathbf{X}_2$  contains additional explanatory variables of which we are less certain. Using the terminology of Danilov and Magnus (2004), the  $k_1$  columns of  $\mathbf{X}_1$  are called focus regressors and the  $k_2$  columns of  $\mathbf{X}_2$  are called auxiliary regressors.

Our primary interest is the estimation of the vector of focus parameters  $\boldsymbol{\beta}_1$ , while  $\boldsymbol{\beta}_2$  is treated as a vector of nuisance parameters. By the properties of partitioned inverses, the unrestricted

---

<sup>1</sup> Updated versions of the Matlab commands for BMA and WALS are can be downloaded free of charge from the website [www.janmagnus.nl/items/BMA.pdf](http://www.janmagnus.nl/items/BMA.pdf).

ordinary least squares (OLS) estimators of  $\beta_1$  and  $\beta_2$  are given by

$$\widehat{\beta}_{1u} = \widehat{\beta}_{1r} - \mathbf{Q}\widehat{\beta}_{2u}, \quad \widehat{\beta}_{2u} = (\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{M}_1 \mathbf{y},$$

where  $\widehat{\beta}_{1r} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{y}$  is the restricted OLS estimator from a regression of  $\mathbf{y}$  on  $\mathbf{X}_1$  (with  $\beta_2$  restricted to zero),  $\mathbf{Q} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2$  is the multivariate OLS estimator from a regression of  $\mathbf{X}_2$  on  $\mathbf{X}_1$ , and  $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top$  is a symmetric and idempotent matrix. Within this framework, model uncertainty arises because different subsets of auxiliary regressors could be excluded from  $\mathbf{X}_2$  to improve, in the mean squared error (MSE) sense, the unrestricted OLS estimator  $\widehat{\beta}_{1u}$  of  $\beta_1$ . It is a basic results from least squares theory that by restricting some elements of  $\beta_2$  to zero we can indeed obtain an estimator of  $\beta_1$  which is subject to omitted variable bias but is also more precise than the unrestricted OLS estimator  $\widehat{\beta}_{1u}$ . The choice of excluding different subsets of auxiliary regressors is therefore motivated by a trade-off between bias and precision in the estimators of the focus regression parameters.

Since model uncertainty is confined to the  $k_2$  variables of  $\mathbf{X}_2$ , the number of possible models to be considered is  $I = 2^{k_2}$ . In what follows, we denote by  $\mathcal{M}_i$  the  $i$ th model in the model space which is obtained by including only a subset of  $k_{2i}$  (with  $0 \leq k_{2i} \leq k_2$ ) auxiliary regressors. Model  $\mathcal{M}_i$  is represented as follows

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_{2i} \beta_{2i} + \epsilon_i, \quad i = 1, \dots, I, \quad (2)$$

where  $\mathbf{X}_{2i}$  is an  $n \times k_{2i}$  matrix of observations on the included subset of  $k_{2i}$  auxiliary regressors,  $\beta_{2i}$  is the corresponding subvector of auxiliary parameters, and  $\epsilon_i$  is the new vector of disturbances after excluding  $k_2 - k_{2i}$  auxiliary regressors.

### 3 Model averaging estimators

The basic idea of model averaging estimators is that one first estimates the parameters of interest conditional on each model in the model space, and then computes the unconditional estimate as a weighted average of these conditional estimates. A model averaging estimate of  $\beta_1$  is given by

$$\widehat{\beta}_1 = \sum_{i=1}^I \lambda_i \widehat{\beta}_{1i}, \quad (3)$$

where the  $\lambda_i$  are non-negative random weights that add up to one, and  $\widehat{\beta}_{1i}$  is the estimate of  $\beta_1$  obtained by conditioning on model  $\mathcal{M}_i$ . Below, we discuss two model averaging estimators.

### 3.1 Bayesian model averaging

The Bayesian Model Averaging (BMA) estimator developed in Magnus et al. (2010) generalizes the framework used in standard BMA estimation by introducing the distinction between focus and auxiliary regressors. Like other Bayesian estimators, this estimator combines prior beliefs on the unknown elements of the model with the additional information coming from the data. Its key ingredients are the sample likelihood function, the prior distributions on the regression parameters of model  $\mathcal{M}_i$ , and the prior distributions on the model space.

If we assume that  $\mathcal{M}_i$  is the true model, then the sample likelihood function implied by model (2) can be written as

$$p(\mathbf{y} | \boldsymbol{\beta}_1, \boldsymbol{\beta}_{2i}, \sigma^2, \mathcal{M}_i) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{\boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}_i}{2\sigma^2}\right). \quad (4)$$

Prior beliefs on the regression parameters of model  $\mathcal{M}_i$  are introduced by imposing conventional non-informative priors on the focus parameters  $\boldsymbol{\beta}_1$  and the error variance  $\sigma^2$ , plus an informative Gaussian prior on the auxiliary parameters  $\boldsymbol{\beta}_{2i}$ . This leads to a conditional joint prior distribution of the form

$$p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2 | \mathcal{M}_i) \propto (\sigma^2)^{(k_{2i}+2)/2} \exp\left(-\frac{\boldsymbol{\beta}_{2i}^\top \mathbf{V}_{0i}^{-1} \boldsymbol{\beta}_{2i}}{2\sigma^2}\right), \quad (5)$$

where  $\mathbf{V}_{0i}$  is the variance-covariance matrix of the prior distribution of  $\boldsymbol{\beta}_{2i}$  which takes the standard form proposed by Zellner (1986) and Fernández et al. (2001)

$$\mathbf{V}_{0i}^{-1} = g \mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i},$$

and  $g = 1/\max(n, k_2^2)$  is a constant scalar for each model  $\mathcal{M}_i$ .

In Bayesian inference, we would like to combine the likelihood function (4) with the conditional joint prior distribution (5) to obtain the conditional posterior distribution  $p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2i}, \sigma^2 | \mathbf{y}, \mathcal{M}_i)$ . As argued by Magnus et al. (2010), this task is complicated by the fact that the assumed prior distribution involves partially proper and partially improper priors. To overcome this problem, they use a more general proper prior that admits the improper prior in (5) as a limiting case. After computing the conditional posterior distribution on the basis of this more general prior and specializing the results to the assumed prior, Magnus et al. (2010) show that the conditional estimates of  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_{2i}$  under model  $\mathcal{M}_i$  are given by

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{1i} &= E(\boldsymbol{\beta}_1 | \mathbf{y}, \mathcal{M}_i) = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{X}_{2i} \widehat{\boldsymbol{\beta}}_{2i}), \\ \widehat{\boldsymbol{\beta}}_{2i} &= E(\boldsymbol{\beta}_{2i} | \mathbf{y}, \mathcal{M}_i) = (1 + g)^{-1} (\mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i})^{-1} \mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{y}. \end{aligned} \quad (6)$$

Provided that  $n > k_1 + 2$ , the elements of the variance-covariance matrix are given by

$$\begin{aligned}\widehat{\mathbf{V}}_{1i} &= \text{Var}(\boldsymbol{\beta}_1 | \mathbf{y}, \mathcal{M}_i) = s_i^2 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} + \mathbf{Q}_i \widehat{\mathbf{V}}_{2i} \mathbf{Q}_i^\top, \\ \widehat{\mathbf{V}}_{2i} &= \text{Var}(\boldsymbol{\beta}_{2i} | \mathbf{y}, \mathcal{M}_i) = s_i^2 (1 + g)^{-1} (\mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i})^{-1}, \\ \widehat{\mathbf{V}}_{12i} &= \text{Cov}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2i} | \mathbf{y}, \mathcal{M}_i) = -\mathbf{Q}_i \widehat{\mathbf{V}}_{2i},\end{aligned}\tag{7}$$

where  $s_i^2 = (\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y}) / (n - k_1 - 2)$  is the estimate of  $\sigma^2$  under model  $\mathcal{M}_i$ ,  $\mathbf{Q}_i = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_{2i}$  is the multivariate OLS estimator from a regression of  $\mathbf{X}_{2i}$  on  $\mathbf{X}_1$ , and

$$\mathbf{A}_i = \frac{g}{1+g} \mathbf{M}_1 + \frac{1}{1+g} \left[ \mathbf{M}_1 - \mathbf{M}_1 \mathbf{X}_{2i} (\mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i})^{-1} \mathbf{X}_{2i}^\top \mathbf{M}_1 \right].$$

Prior beliefs on the model space are introduced by assuming that each model is weighted by its posterior probability

$$\lambda_i = p(\mathcal{M}_i | \mathbf{y}) = \frac{p(\mathcal{M}_i) p(\mathbf{y} | \mathcal{M}_i)}{\sum_{j=1}^I p(\mathcal{M}_j) p(\mathbf{y} | \mathcal{M}_j)},$$

where  $p(\mathcal{M}_i)$  is the prior probability of model  $\mathcal{M}_i$  and  $p(\mathbf{y} | \mathcal{M}_i)$  is the marginal likelihood of  $\mathbf{y}$  given model  $\mathcal{M}_i$ . By assigning equal prior probabilities  $p(\mathcal{M}_i) = 2^{-k_2}$  to each model and exploiting the above assumptions on the joint prior distribution, one can show that

$$\lambda_i = p(\mathbf{y} | \mathcal{M}_i) = c \left( \frac{g}{1+g} \right)^{k_{2i}/2} (\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y})^{-(n-k_1)/2},\tag{8}$$

where  $c$  is a normalizing constant chosen to guarantee that the  $\lambda_i$  add up to one (see Section 6).

Given the conditional estimates  $\widehat{\boldsymbol{\beta}}_{1i}$  and  $\widehat{\boldsymbol{\beta}}_{2i}$  of the regression parameters of model  $\mathcal{M}_i$  and the model weights  $\lambda_i$ , the unconditional BMA estimates of  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are computed as follows

$$\widehat{\boldsymbol{\beta}}_1 = \text{E}(\boldsymbol{\beta}_1 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \widehat{\boldsymbol{\beta}}_{1i}, \quad \widehat{\boldsymbol{\beta}}_2 = \text{E}(\boldsymbol{\beta}_2 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \mathbf{T}_i \widehat{\boldsymbol{\beta}}_{2i},\tag{9}$$

where the  $\mathbf{T}_i$  are  $k_2 \times k_{2i}$  matrices defined by  $\mathbf{T}_i^\top = (\mathbf{I}_{k_{2i}}, \mathbf{0})$ , or a column-permutation thereof, that transform the conditional estimates  $\widehat{\boldsymbol{\beta}}_{2i}$  in  $k_2 \times 1$  vectors by setting to zero the elements of  $\boldsymbol{\beta}_2$  which are excluded from model  $\mathcal{M}_i$ . The elements of the posterior variance-covariance matrix are given by

$$\begin{aligned}\text{Var}(\widehat{\boldsymbol{\beta}}_1 | \mathbf{y}) &= \sum_{i=1}^I \lambda_i \left( \widehat{\mathbf{V}}_{1i} + \widehat{\boldsymbol{\beta}}_{1i} \widehat{\boldsymbol{\beta}}_{1i}^\top \right) - \widehat{\boldsymbol{\beta}}_1 \widehat{\boldsymbol{\beta}}_1^\top, \\ \text{Var}(\widehat{\boldsymbol{\beta}}_2 | \mathbf{y}) &= \sum_{i=1}^I \lambda_i \mathbf{T}_i \left( \widehat{\mathbf{V}}_{2i} + \widehat{\boldsymbol{\beta}}_{2i} \widehat{\boldsymbol{\beta}}_{2i}^\top \right) \mathbf{T}_i^\top - \widehat{\boldsymbol{\beta}}_2 \widehat{\boldsymbol{\beta}}_2^\top, \\ \text{Cov}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2 | \mathbf{y}) &= \sum_{i=1}^I \lambda_i \left( \widehat{\mathbf{V}}_{12i} + \widehat{\boldsymbol{\beta}}_{1i} \widehat{\boldsymbol{\beta}}_{2i}^\top \right) \mathbf{T}_i^\top - \widehat{\boldsymbol{\beta}}_1 \widehat{\boldsymbol{\beta}}_2^\top.\end{aligned}\tag{10}$$

Notice that, unlike pretest estimators, these variances take into account the uncertainty due to both the parameter estimation step and the model selection step. The elements of the variance-covariance matrix consist indeed of two components: the weighted average of the conditional variance-covariance matrices in each model and the weighted variance-covariance matrix of the conditional estimates across all possible models.

Although BMA is a widely used technique, it suffers from two major problems. First, the computational burden required to obtain an exact BMA estimate is proportional to the dimension of the model space  $I = 2^{k_2}$ . Thus, unless the number of auxiliary regressors is small or moderate, this computational burden can be substantial. Second, the choice of the prior distribution on  $\beta_2$  may not be attractive in situations where no prior information is indeed available. Furthermore, the chosen priors imply that the risk of the BMA estimator is unbounded and that our prior beliefs on the same parameters vary across models.

### 3.2 Weighted average least squares

Weighted Average Least Squares (WALS) is an alternative model averaging technique that was originally introduced by Magnus and Durbin (1999) and Danilov and Magnus (2004) to investigate the statistical properties of pretest estimators.

Unlike BMA, WALS relies on preliminary orthogonal transformations of the auxiliary regressors and their parameters which greatly reduce the computational burden of this model averaging estimator and allow exploiting prior distributions corresponding to a more transparent concept of ignorance about the role of the auxiliary regressors. The first step of WALS consists of computing an orthogonal  $k_2 \times k_2$  matrix  $\mathbf{P}$  and a diagonal  $k_2 \times k_2$  matrix  $\mathbf{\Lambda}$  such that  $\mathbf{P}^\top \mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2 \mathbf{P} = \mathbf{\Lambda}$ . These matrices are then used to define  $\mathbf{Z}_2 = \mathbf{X}_2 \mathbf{P} \mathbf{\Lambda}^{-1/2}$  and  $\gamma_2 = \mathbf{\Lambda}^{1/2} \mathbf{P}^\top \beta_2$  such that  $\mathbf{Z}_2^\top \mathbf{M}_1 \mathbf{Z}_2 = \mathbf{I}_{K_2}$  and  $\mathbf{Z}_2 \gamma_2 = \mathbf{X}_2 \beta_2$ . Notice that the original vector of auxiliary parameters  $\beta_2$  can be always recovered from  $\beta_2 = \mathbf{P} \mathbf{\Lambda}^{-1/2} \gamma_2$ .

After applying these orthogonal transformations to model (1), the unrestricted OLS estimators of  $\beta_1$  and  $\gamma_2$  from a regression of  $\mathbf{y}$  on  $\mathbf{X}_1$  and  $\mathbf{Z}_2$  are given by

$$\hat{\beta}_{1u} = \hat{\beta}_{1r} - \mathbf{R} \hat{\gamma}_{2u}, \quad \hat{\gamma}_{2u} = \mathbf{Z}_2^\top \mathbf{M}_1 \mathbf{y},$$

where  $\mathbf{R} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{Z}_2$  is the multivariate OLS estimator from a regression of  $\mathbf{Z}_2$  on  $\mathbf{X}_1$ . If we also define the  $k_2 \times (k_2 - k_{2i})$  selection matrix  $\mathbf{S}_i$  by  $\mathbf{S}_i^\top = (\mathbf{I}_{k_2 - k_{2i}}, \mathbf{0})$ , or a column-permutation thereof, so that  $\mathbf{S}_i$  captures the restrictions placed on  $\gamma_2$  under model  $\mathcal{M}_i$ , then the restricted OLS



estimators of  $\beta_1$  and  $\gamma_{2i}$  are given by

$$\widehat{\beta}_{1i} = \widehat{\beta}_{1r} - \mathbf{R}\mathbf{W}_i\widehat{\gamma}_{2u}, \quad \widehat{\gamma}_{2i} = \mathbf{W}_i\widehat{\gamma}_{2u},$$

where  $\mathbf{W}_i = \mathbf{I}_{k_2} - \mathbf{S}_i\mathbf{S}_i^\top$  is a diagonal  $k_2 \times k_2$  matrix whose  $j$ th diagonal element is equal to zero if  $\gamma_{2j}$  is restricted to zero and is equal to one otherwise.

The key advantage of these transformations lies in the fact that  $\widehat{\gamma}_{2u} \sim N_{k_2}(\gamma_2, \sigma^2\mathbf{I}_{k_2})$ . This result has a number of implications on the computational aspects and the statistical properties of the WALS estimator. First, under some minimal regularity conditions on the model weights  $\lambda_i$ , the WALS estimator of  $\beta_1$  is of the form

$$\widetilde{\beta}_1 = \sum_{i=1}^I \lambda_i \widehat{\beta}_{1i} = \widehat{\beta}_{1r} - \mathbf{R}\mathbf{W}\widehat{\gamma}_2,$$

where  $\mathbf{W} = \sum_{i=1}^I \lambda_i \mathbf{W}_i$  is a  $k_2 \times k_2$  diagonal random matrix (because the  $\lambda_i$  are random). This shows that, even if the model space contains  $2^{k_2}$  models, the computational burden of the WALS estimator  $\widetilde{\beta}_1$  is of the order  $k_2$  because we need only consider the diagonal elements of  $\mathbf{W}$ , that is  $k_2$  linear combinations of the model weights  $\lambda_i$ .

Second, the equivalence theorem proved in Danilov and Magnus (2004) implies that the MSE of the WALS estimator  $\widetilde{\beta}_1$  of  $\beta_1$  is crucially related to the MSE of the less complicated shrinkage estimator  $\mathbf{W}\widehat{\gamma}_2$  of  $\gamma_2$ ,

$$\text{MSE}(\widetilde{\beta}_1) = \sigma^2 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} + \mathbf{R} \text{MSE}(\mathbf{W}\widehat{\gamma}_2) \mathbf{R}^\top.$$

Thus, if we can find the diagonal elements of  $\mathbf{W}$  such that the shrinkage estimator  $\mathbf{W}\widehat{\gamma}_2$  is an optimal estimator of  $\gamma_2$ , then the same estimator will also provide the optimal WALS estimator  $\widetilde{\beta}_1$  of  $\beta_1$ .

Third, since the  $k_2$  components of  $\gamma_2$  are independent, they can be estimated separately by exploiting the information that  $\widehat{\gamma}_{2j} \sim N(\gamma_{2j}, \sigma^2)$ . In Magnus et al. (2010), this problem is addressed using a Laplace estimator  $\widehat{\eta}_j$  for the theoretical  $t$ -ratio  $\eta_j = \gamma_{2j}/\sigma$ . This choice is motivated by the results in Magnus (2002) who shows that  $\widehat{\eta}_j$  is admissible, has bounded risk, has good properties around  $|\eta| = 1$  and is nearly optimal in terms of a well-defined regret criterion.<sup>2</sup> Furthermore, this Bayesian estimator is based on a Laplace prior distribution

$$\pi(\eta_j; c) = \frac{c}{2} \exp(-c|\eta_j|), \quad (11)$$

---

<sup>2</sup> Notice that, in estimating  $\eta_j$ , the unknown parameter  $\sigma^2$  is replaced by the unbiased estimator  $s^2$  obtained from the unrestricted model. The results in Danilov (2005) show that this approximation has only marginal effects on the statistical properties of this estimator.

with  $c = \log 2$  to satisfy the property of neutrality (i.e. the prior median of  $\eta_j$  is zero and the prior median of  $\eta_j^2$  is one) which reflects our notion of ignorance in situations where we do not know whether the  $t$ -ratio  $\eta_j$  is larger or smaller than one in absolute value. The WALs estimator proposed by Einmahl et al. (2011) uses instead an estimator  $\tilde{\eta}_j$  of  $\eta_j$  based on the Subbotin density

$$\pi(\eta_j; q, c) = \frac{qc^{1/q}}{2\Gamma(1/q)} \exp(-c|\eta_j|^q), \quad (12)$$

with  $c > 0$  and  $q > 0$ .<sup>3</sup> This prior allows obtaining a class of estimators  $\tilde{\eta}_j(q, c)$  with better properties than the Laplace estimator  $\hat{\eta}_j$ , especially when  $\eta_j$  is large. As for the choice of the parameters  $q$  and  $c$ , Einmahl et al. (2011) show that  $q$  must belong to the interval  $(0, 1)$  in order to obtain a well-behaved estimator of  $\eta_j$ . Given  $q = \bar{q}$ , the parameter  $c$  can be chosen implicitly by solving the non-linear equation

$$\int_0^1 \pi(\eta; \bar{q}, c) d\eta = \frac{1}{4}, \quad (13)$$

to satisfy neutrality. Figure 1 plots a neutral Subbotin density with free parameter  $q = 0.5$  together with a Laplace density ( $q = 1$ ) and a Gaussian density ( $q = 2$ ). We can see that a value of  $q < 1$  corresponds to a density which is less flat in the interval  $(0, 1)$  and has thicker tails. For empirical applications, Einmahl et al. (2011) recommend using a Subbotin prior with  $q = 0.5$ .

Let us denote by  $\bar{\boldsymbol{\eta}}$  the Laplace or the Subbotin estimator of  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{k_2})$ . Magnus et al. (2010) show that the WALs estimators of the regression parameters  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are given by

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_1 &= (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{X}_2 \tilde{\boldsymbol{\beta}}_2), \\ \tilde{\boldsymbol{\beta}}_2 &= s \mathbf{P} \boldsymbol{\Lambda}^{-1/2} \bar{\boldsymbol{\eta}}, \end{aligned} \quad (14)$$

and the elements of their variance-covariance matrix are

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\beta}}_1) &= s^2 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} + \mathbf{Q} \text{Var}(\tilde{\boldsymbol{\beta}}_2) \mathbf{Q}^\top, \\ \text{Var}(\tilde{\boldsymbol{\beta}}_2) &= s^2 \mathbf{P} \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\Omega} \boldsymbol{\Lambda}^{-1/2} \mathbf{P}, \\ \text{Cov}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2) &= -\mathbf{Q} \text{Var}(\tilde{\boldsymbol{\beta}}_2), \end{aligned} \quad (15)$$

where  $\mathbf{Q} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2$  and  $\boldsymbol{\Omega}$  is the diagonal variance-covariance matrix of  $\bar{\boldsymbol{\eta}}$ . It is worth noticing that this model averaging technique can also be generalized to non-spherical errors (see Magnus et al. 2011). Thus, the assumption of homoskedastic and serially uncorrelated regression errors is not crucial for WALs.

---

<sup>3</sup> The Gaussian and the Laplace densities can be obtained as special cases of the Subbotin density by setting ( $q = 2, c = 1/2$ ) and ( $q = 1, c = \log 2$ ) respectively.

## 4 Equivariance

An estimator may be equivariant to a certain transformation or not.<sup>4</sup> If the transformation is considered to be ‘trivial’, then we prefer the estimator to be equivariant, that is, not to change other than in a trivial fashion. For example, in the basic regression model

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

with  $E(\mathbf{u}) = \mathbf{0}$  and  $\text{Var}(\mathbf{u}) = \sigma^2\mathbf{I}_n$ , we generally do not want the *ordering* of the columns in  $\mathbf{X}$  to influence the outcome. If we find  $\hat{\beta}_2 = 2$  and  $\hat{\beta}_3 = 3$ , and then estimate again but now interchanging  $\mathbf{x}_2$  and  $\mathbf{x}_3$ , then we expect to find  $\hat{\beta}_2 = 3$  and  $\hat{\beta}_3 = 2$  in the new ordering. Hence, the estimates have changed but in a trivial fashion. It is also possible that the estimates change in a non-trivial fashion. An example is given by sequential model selection procedures based on a hierarchical order of the regressors. In general, BMA and WALS estimators are equivariant with respect to the ordering of focus and auxiliary regressors. However, if we interchange a focus regressor with an auxiliary regressor, then estimates change in a non-trivial fashion because such a transformation corresponds to a different model specification.

Another common transformation is *shift*. If we consider, instead of  $\boldsymbol{\beta}$ , a translation  $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ , then the regression equation can be written as

$$\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0 = \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{u},$$

and the quadratic estimator  $\mathbf{y}^\top \mathbf{A}\mathbf{y}$  of  $\sigma^2$  is shift-equivariant if

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)^\top \mathbf{A}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0) = \mathbf{y}^\top \mathbf{A}\mathbf{y}, \quad \text{for all } \boldsymbol{\beta}_0.$$

This is the case if and only if we restrict  $\mathbf{A}$  to satisfy  $\mathbf{A}\mathbf{X} = \mathbf{0}$ . On the other hand, if we require that the estimator  $\mathbf{y}^\top \mathbf{A}\mathbf{y}$  has minimum variance in the class of unbiased estimators, then we obtain the conditions  $\mathbf{X}^\top \mathbf{A}\mathbf{X} = \mathbf{0}$  and  $\text{tr}(\mathbf{A}) = 1$ ; see Magnus and Neudecker (1988, Chapter 14, Sections 1–8). These are not the same conditions, and hence we obtain different estimators. This shows that two reasonable requirements (unbiasedness and shift-equivariance) may not be possible at the same time.

A special case of shift is *centering*. If there is no constant term in the regression and we center the regressors, then the OLS estimates are affected, and the same is true for BMA and WALS. If there is a constant term among the focus variables in the regression and we center the regressors,

---

<sup>4</sup> For a formal treatment of the principle of equivariance see Lehmann and Casella (1998, Chapter 3).

then neither  $\mathbf{M}_1$  nor  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$  is affected, so that BMA and WALS estimates are both equivariant to centering. The reason is simple. Suppose that the first column of  $\mathbf{X}_1$  is  $\mathbf{1}$ , the vector of ones. After centering, we can write the centered matrix as  $\mathbf{X}_1^c = \mathbf{X}_1 \mathbf{E}$ , where  $\mathbf{E}$  is the non-singular  $k_1 \times k_1$  matrix

$$\mathbf{E} = \begin{pmatrix} 1 & -\boldsymbol{\mu}_1^\top \\ \mathbf{0} & \mathbf{I}_{k_1-1} \end{pmatrix},$$

and  $\boldsymbol{\mu}_1$  is a  $(k_1 - 1) \times 1$  vector containing the sample means of the focus regressors (except the constant term). Hence,

$$\begin{aligned} \mathbf{M}_1^c &= \mathbf{I}_n - \mathbf{X}_1^c (\mathbf{X}_1^{c\top} \mathbf{X}_1^c)^{-1} \mathbf{X}_1^{c\top} \\ &= \mathbf{I}_n - (\mathbf{X}_1 \mathbf{E}) ((\mathbf{X}_1 \mathbf{E})^\top (\mathbf{X}_1 \mathbf{E}))^{-1} (\mathbf{X}_1 \mathbf{E})^\top \\ &= \mathbf{I}_n - \mathbf{X}_1 \mathbf{E} (\mathbf{E}^\top \mathbf{X}_1^\top \mathbf{X}_1 \mathbf{E})^{-1} \mathbf{E}^\top \mathbf{X}_1^\top \\ &= \mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top = \mathbf{M}_1. \end{aligned}$$

Also, if  $\boldsymbol{\mu}_2$  is a  $k_2 \times 1$  vector containing the sample means of the auxiliary regressors, then

$$\mathbf{M}_1 \mathbf{X}_2^c = \mathbf{M}_1 (\mathbf{X}_2 - \mathbf{1} \boldsymbol{\mu}_2^\top) = \mathbf{M}_1 \mathbf{X}_2,$$

because  $\mathbf{M}_1 \mathbf{1} = \mathbf{0}$ . This shows that  $\mathbf{M}_1$  and  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$  are both invariant to centering.

A third ‘trivial’ transformation is *scaling*. If we measure each component of one regressor, say  $\mathbf{x}_2$ , in kilos rather than in grams, then we expect nothing to change other than that  $\hat{\beta}_2$  is multiplied by 1000. In a standard (non-Bayesian) context the OLS estimator is scale-independent, but in a Bayesian context, this is so only if data and priors are scaled correspondingly. This is automatically achieved in BMA, but not in WALS. Scaling the focus regressors  $\mathbf{X}_1$  will have no effect on the WALS estimates, but scaling the auxiliary regressors  $\mathbf{X}_2$  will have an effect, unless  $k_2 = 1$ . The reason lies in the semi-orthogonalization, which gives us great benefits, but at the same time make the estimator scale-dependent, because the orthogonal matrix  $\mathbf{P}$  and the diagonal matrix  $\boldsymbol{\Lambda}$  will depend on the scaling in a non-trivial (non-linear) fashion. This property of WALS has not been noticed before, so we emphasize it here and propose a simple remedy. Specifically, we scale the regressors in  $\mathbf{X}_1$  and  $\mathbf{X}_2$  such that the diagonal elements of the matrices  $\mathbf{X}_1^\top \mathbf{X}_1$  and  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$  are all one. Notice that this also stabilizes both matrices so that inversion and eigenvalue routines are numerically more stable. The effect of the scaling in  $\mathbf{X}_1$  is only for numerical stability, but the scaling in  $\mathbf{X}_2$  has two effects: numerical stability and scale-independence.

## 5 Stata commands

The new Stata commands `bma` and `wals` provide BMA and WALS estimates, respectively, of linear regression models with uncertainty about the choice of the explanatory variables. The syntax of these commands is as follows:

```
bma depvar [varlist] [if] [in] , auxiliary(varlist) [ nodots notable noconstant ]  
wals depvar [varlist] [if] [in] , auxiliary(varlist) [ q(#) intpoints(#) eps(#) iterate(#) noconstant ]
```

where `depvar` is the dependent variable, `varlist` is the optional list of focus regressors (including the constant term, if any) which are included with certainty in the model, and `auxiliary` is the required list of auxiliary regressors of which we are less certain. Both commands are programmed in Mata on the basis of the original Matlab commands associated with Magnus et al. (2010). The earliest version of Stata that can be used to run our commands is `version 11.1`. Factor variables, time-series operators and weights are not allowed.

### 5.1 Options for BMA

`nodots` suppresses the display of the dots to track the progress of `bma` estimation. Dots are displayed only if the model space consists of more than 128 models (i.e. at least 7 auxiliary regressors).

One dot means that 1% of the models in the model space has been estimated.

`notable` suppresses the display of the table of results.

`noconstant` specifies that constant term must excluded from the model. By default the constant term is included and the corresponding vector of ones is treated as a focus regressor.

### 5.2 Options for WALS

`q(#)` specifies the free parameter  $0 < q \leq 1$  of a Subbotin prior distribution under neutrality. The default is  $q = 1$  which corresponds to a neutral Laplace prior, while any real value of  $q$  in the interval  $(0, 1)$  corresponds to a neutral Subbotin prior.

`intpoints(#)` defines the number of data points used by the built-in Stata command `integ` when approximating numerically the integral involved in the non-linear equation for the constrained parameter  $c$  of a Subbotin density under neutrality. The default uses 10000 data points. Notice that, for  $q = 1$  and  $q = 0.5$ , this option is ineffective because the solution of the constrained parameter  $c$  is determined analytically. Similar considerations hold for the options `eps(#)` and `iterate(#)`.

`eps(#)` specifies the convergence criterion used by the built-in Stata command `n1` when solving the non-linear equation for the constrained parameter  $c$  of a Subbotin density under neutrality.

The default is `1e-8`.

`iterate(#)` specifies the maximum number of iterations used by the built-in Stata command `n1` when solving the non-linear equation for the constrained parameter  $c$  of a Subbotin density under neutrality. The default is `16000`.

`noconstant` same as the `noconstant` option of `bma` command.

## 6 Additional remarks

1. The Stata command `bma` improves the original Matlab command for BMA estimation in two respects. First, our command uses a more stable normalization of the model weights to avoid numerical problems in the computation of BMA estimates. In the original Matlab command, the model weights are normalized with respect to the weight of the restricted model by imposing that  $\lambda_i = \lambda_i^*/\Lambda^*$ ,  $\Lambda^* = \sum_i \lambda_i^*$ ,  $\lambda_1^* = 1$  and

$$\lambda_i^* = \left( \frac{g}{1+g} \right)^{k_{2i}/2} \left( \frac{\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y}}{\mathbf{y}^\top \mathbf{M}_1 \mathbf{y}} \right)^{-(n-k_1)/2}, \quad i = 2, \dots, I.$$

Notice that, if the sample size is large, this normalization may lead to numerically too large model weights because  $(\mathbf{y}^\top \mathbf{M}_1 \mathbf{y}) > (\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y})$  for each  $i = 2, \dots, I$ . In our `bma` command, the model weights are instead scaled with respect to the weight of the unrestricted model by imposing that  $\lambda_I^* = 1$  and

$$\lambda_i^* = \left( \frac{g}{1+g} \right)^{k_{2i}/2} \left( \frac{\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y}}{\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_I \mathbf{M}_1 \mathbf{y}} \right)^{-(n-k_1)/2}, \quad i = 1, \dots, I-1.$$

Given that  $(\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_I \mathbf{M}_1 \mathbf{y}) < (\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y})$  for each  $i = 1, \dots, I-1$ , this normalization guarantees that the  $\lambda_i^*$  are always bounded in the  $(0, 1)$  interval. Second, whenever the sample size is moderately large (say  $n > 100$ ), our command is considerably faster and requires much less memory than the corresponding Matlab command because it avoids computing matrices of order  $n \times n$ .

2. The Stata command `wals` improves the original Matlab command for WALS estimation in two respects. First, it introduces a preliminary scaling of the regressors in  $\mathbf{X}_1$  and  $\mathbf{X}_2$  to ensure scale-equivariance and greater accuracy of the WALS estimates. Second, it allows specifying neutral Subbotin priors with any real value of  $q$  in the interval  $(0, 1)$  instead of a

list of nine focal values  $q = (.1, .2, \dots, .9)$ . Our `wals` Stata command also differs from the Matlab command because moments of the Subbotin density are calculated by Gauss-Laguerre quadrature methods with 100 data points instead of high-order global adaptive quadrature methods.<sup>5</sup>

3. The Stata commands `bma` and `wals` also improve the corresponding Matlab commands because they use listwise deletion of missing values to deal with problems of missing data in the dependent and the explanatory variables, they do not require any specific ordering of focus and auxiliary regressors within the data set, and they compute estimated covariances between focus and auxiliary parameters.
4. The Stata commands `bma` and `wals` differ from other Stata estimation commands because they do not provide  $p$ -values of the  $t$ -ratios for testing the significance of the estimated regression parameters and their confidence intervals. The Bayesian counterparts of these quantities cannot be easily computed because these estimators are biased and their distributions are not Gaussian. On the other hand, a regressor may be considered to be robustly correlated with the dependent variable if the corresponding  $t$ -ratio is greater than one in absolute value, in which case the MSE of the unrestricted OLS estimator is lower than MSE of the restricted OLS estimator (see Magnus 2002). On the basis of this criterion, our commands provide two-standard error bands of the estimated regression parameters.

## 7 Example

This section uses the growth data analyzed by Magnus et al. (2010) and Einmahl et al. (2011) for illustrating our `bma` and `wals` commands, validating their estimation results and investigating equivariance of the BMA and WALS estimators to shift and scale transformations of the explanatory variables.<sup>6</sup> Data constitute a cross section of the average growth rate of the per-capita GDP between 1960–1996 for 74 countries worldwide.

```
. use Data_MPP_small, clear
. describe
Contains data from Data_MPP_small.dta
  obs:          74
  vars:          11                               30 May 2011 08:55
  size:         4,514 (99.9% of memory free)
```

---

	storage	display	value
--	---------	---------	-------

<sup>5</sup> A description of these alternative quadrature methods can be found in Cheney and Kincaid (2008).

<sup>6</sup> Data can be downloaded from the website [www.janmagnus.nl/items/BMA.pdf](http://www.janmagnus.nl/items/BMA.pdf).

variable name	type	format	label	variable label
country	str17	%17s		Country
growth	float	%9.0g		Growth GDP per capita 1960-1996
gdp60	float	%9.0g		Log of GDP per capita 1960
equipinv	float	%9.0g		Real equipment investment/GDP 1960-1985
confuc	float	%9.0g		Fraction of Confucian population
school60	float	%9.0g		Enrollm. rate primary school 1960
life60	float	%9.0g		Life exp. at age zero 1960
law	float	%9.0g		Rule of law index
tropics	float	%9.0g		Fraction tropical area
avelf	float	%9.0g		Ethnolinguistic fragmentation
dpop	float	%9.0g		Population growth rate 1960-1990

Sorted by:

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
country	0				
growth	74	.01985	.0186466	-.0318	.0691
gdp60	74	7.525295	.8612332	5.549076	9.199785
equipinv	74	.0432353	.0344413	.00135	.1482
confuc	74	.0185135	.0862726	0	.6
school60	74	.7806757	.2555753	.07	1
life60	74	56.06757	11.56578	36.1	73.4
law	74	.5518018	.3332485	0	1
tropics	74	.5480932	.4709772	0	1
avelf	74	.2984383	.2797776	0	.8722529
dpop	74	.0205902	.0099794	.0023685	.0356537

Magnus et al. (2010) and Einmahl et al (2011) provide BMA and WALS estimates of different model specifications to test the implications of alternative growth theories. Here, for simplicity, we focus on Set-up 1 of their Model 1 which allows testing the neoclassical growth theory against the new growth theories of institutions, geography, fractionalization, and religion. The outcome variable of interest is `growth`, the subset of focus regressors includes the constant term and five ‘Solow’ determinants derived from the neoclassical growth theory, while the subset of auxiliary regressors includes four growth determinants derived from the other theories.

```
. local y "growth"
. local X1 "gdp60 equipinv school60 life60 dpop"
. local X2 "law tropics avelf confuc"
```

The BMA estimates of this growth regression model are given by

```
. bma `y' `X1' , aux(`X2')
BMA estimates
```

	Coef.	Std. Err.	t	pip	[2-Std. Err. Bands]
growth					
_cons	.0492403	.0229036	2.15	1.00	.0263367 .0721439
gdp60	-.0138652	.0034982	-3.96	1.00	-.0173633 -.010367
equipinv	.1643892	.0614866	2.67	1.00	.1029026 .2258758

Number of obs = 74  
k1 = 6  
k2 = 4



school60	.0160304	.0101594	1.58	1.00	.005871	.0261898
life60	.0008443	.0003635	2.32	1.00	.0004809	.0012078
dpop	.1654229	.2769833	0.60	1.00	-.1115604	.4424063
law	.01089	.009329	1.17	0.68	.001561	.020219
tropics	-.0035352	.0047221	-0.75	0.45	-.0082574	.0011869
avelf	-.0020815	.0046928	-0.44	0.25	-.0067743	.0026114
confuc	.0611861	.0185069	3.31	0.99	.0426791	.079693

The output of the `bma` command provides information on estimated coefficients and their standard errors (i.e. mean and standard deviation of the posterior distribution),  $t$ -ratios, posterior inclusion probabilities (i.e. the posterior probability that a variable is included in the model) and two-standard error bands. Estimation results for the focus and the auxiliary parameters are displayed in the upper and the lower panels of the table respectively. We notice that estimated coefficients and standard errors coincide exactly with those reported in Table 2 of Magnus et al. (2010) under BMA. An auxiliary regressor is considered to be robustly correlated with the outcome if either the  $t$ -ratio on its coefficient is greater than one in absolute value or, equivalently, the corresponding two-standard error band does not include zero. Alternatively, robustness of the auxiliary regressors can be judged on the basis of their posterior inclusion probabilities. As a rough guideline, Raftery (1995) and Masanjala and Papageorgiou (2008) suggest that a posterior inclusion probability of 0.5 corresponds approximately to a  $t$ -ratio of one in absolute value.

Our validation of the estimation results for WALS is carried out in two steps. First, we present the estimates from a fictitious command `walsns` which implements the original WALS procedure without any preliminary scaling of focus and auxiliary regressors. After showing that we can replicate the original WALS estimates, we present the estimates from our `wals` command which introduces a preliminary scaling of the variables in  $\mathbf{X}_1$  and  $\mathbf{X}_2$  such that the diagonal elements of the matrices  $\mathbf{X}_1^\top \mathbf{X}_1$  and  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$  are all one. The estimates from these commands with a neutral Laplace prior are given by

```
. walsns `y' `X1' , aux(`X2')
```

WALS estimates - Laplace prior

Number of obs	=	74
k1	=	6
k2	=	4
q	=	1.0000
c	=	0.6931
kappa	=	4.3

growth	Coef.	Std. Err.	t	[2-Std. Err. Bands]
_cons	.059387	.0220668	2.69	.0373202 .0814538
gdp60	-.0156304	.0032695	-4.78	-.0188999 -.0123608
equipinv	.1555044	.055129	2.82	.1003754 .2106333
school60	.0174593	.0096968	1.80	.0077625 .0271561
life60	.0008557	.000351	2.44	.0005047 .0012067
dpop	.2650796	.2486804	1.07	.0163993 .51376

law	.0146823	.006536	2.25	.0081462	.0212183
tropics	-.0055319	.0037127	-1.49	-.0092446	-.0018193
avelf	-.0053315	.0048132	-1.11	-.0101447	-.0005183
confuc	.0443343	.0163147	2.72	.0280196	.060649

```
. wals `y' `X1' , aux(`X2')
```

```
WALS estimates - Laplace prior          Number of obs =    74
k1                                       =            6
k2                                       =            4
q                                         =    1.0000
c                                         =    0.6931
kappa                                    =    1.3
```

growth	Coef.	Std. Err.	t	[2-Std. Err. Bands]
_cons	.0617514	.0217909	2.83	.0399605 .0835422
gdp60	-.0156501	.0031439	-4.98	-.018794 -.0125062
equipinv	.1582128	.054421	2.91	.1037918 .2126339
school60	.0166758	.009667	1.73	.0070088 .0263429
life60	.0008515	.0003505	2.43	.000501 .0012019
dpop	.2713869	.2425285	1.12	.0288584 .5139153
law	.0134105	.0058037	2.31	.0076069 .0192142
tropics	-.0059973	.0034556	-1.74	-.0094529 -.0025417
avelf	-.0076757	.0050657	-1.52	-.0127414 -.00261
confuc	.046455	.0142765	3.25	.0321785 .0607316

The output of the `wals` command is similar to that of `bma`. The main difference is that WALS does not allow computing the posterior inclusion probabilities because this model averaging technique considers only  $k_2$  linear combinations of the model weights  $\lambda_i$ . We can see that the estimates from the fictitious `walsns` command coincide exactly with those reported in Table 2 of Magnus et al. (2010) under WALS. The estimates from our `wals` command are slightly different because the orthogonal transformations applied in this technique depend on scaling of the auxiliary regressors in a non-linear way. As argued in Section 4, the aim of this preliminary scaling step is twofold: (i) to make the WALS estimator equivariant to scale transformations of the auxiliary regressors and (ii) to improve accuracy of the WALS estimates. As measure of inaccuracy, the output of our command provides the square root of the condition number of the matrix  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$

$$\kappa = \sqrt{\frac{\lambda_{\max}(\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2)}{\lambda_{\min}(\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2)}} \geq 1,$$

where  $\lambda_{\max}$  and  $\lambda_{\min}$  denote, respectively, the maximum and minimum eigenvalues of  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$ . The larger is  $\kappa$  the more ill-conditioned is the matrix  $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$ . In other words, a large value of  $\kappa$  indicates that this matrix is almost singular and the inverse and eigenvalue routines used in the orthogonal transformations of the auxiliary regressors and their parameters can be prone to large numerical errors. Although in the empirical application under examination numerical problems are

not worrisome, we can see that after scaling  $\kappa$  decreases from 4.3 to 1.3.

Before investigating the effects of scale transformations, we show that the estimators considered in this article are equivariant to shift transformations of focus and auxiliary regressors. In what follows, we compare estimates from `bma`, `walsns` and `wals` after centering either the focus regressor `gdp60` or the auxiliary regressor `law` to their sample means.

```
. local method "bma walsns wals"
. use Data_MPP_small, clear
. quietly summarize gdp60
. quietly replace gdp60=gdp60-r(mean)
. foreach m of local method {
2.     quietly `m' `y' `X1' , aux(`X2')
3.     estimates store `m'1
4. }
. use Data_MPP_small, clear
. quietly summarize law
. quietly replace law=law-r(mean)
. foreach m of local method {
2.     quietly `m' `y' `X1' , aux(`X2')
3.     estimates store `m'2
4. }
. estimates table bma1 bma2 walsns1 walsns2 wals1 wals2, ///
>     b(%7.4f) se(%7.4f)
```

Variable	bma1	bma2	walsns1	walsns2	wals1	wals2
_cons	-0.0551	0.0552	-0.0582	0.0675	-0.0560	0.0692
gdp60	0.0195	0.0243	0.0184	0.0228	0.0183	0.0221
	-0.0139	-0.0139	-0.0156	-0.0156	-0.0157	-0.0157
	0.0035	0.0035	0.0033	0.0033	0.0031	0.0031
equipinv	0.1644	0.1644	0.1555	0.1555	0.1582	0.1582
	0.0615	0.0615	0.0551	0.0551	0.0544	0.0544
school60	0.0160	0.0160	0.0175	0.0175	0.0167	0.0167
	0.0102	0.0102	0.0097	0.0097	0.0097	0.0097
life60	0.0008	0.0008	0.0009	0.0009	0.0009	0.0009
	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
dpop	0.1654	0.1654	0.2651	0.2651	0.2714	0.2714
	0.2770	0.2770	0.2487	0.2487	0.2425	0.2425
law	0.0109	0.0109	0.0147	0.0147	0.0134	0.0134
	0.0093	0.0093	0.0065	0.0065	0.0058	0.0058
tropics	-0.0035	-0.0035	-0.0055	-0.0055	-0.0060	-0.0060
	0.0047	0.0047	0.0037	0.0037	0.0035	0.0035
avelf	-0.0021	-0.0021	-0.0053	-0.0053	-0.0077	-0.0077
	0.0047	0.0047	0.0048	0.0048	0.0051	0.0051
confuc	0.0612	0.0612	0.0443	0.0443	0.0465	0.0465
	0.0185	0.0185	0.0163	0.0163	0.0143	0.0143

Legend: b/se

As expected, estimates of the slope coefficients are invariant to centering, while the estimate of the intercept coefficient changes in a trivial fashion. For example, after centering the focus regressor `gdp60`, the new BMA estimate of the intercept coefficient in `bma1` is given by  $.0492403 + 7.525295 * (-.0138652) = -.05509942$ . The effects of scale transformations can be assessed in a similar way. Below, we compare estimates from `bma`, `walsns` and `wals` after dividing either the focus regressor

gdp60 or the auxiliary regressor law by 100.

```
. use Data_MPP_small, clear
. quietly replace gdp60=gdp60/100
. foreach m of local method {
2.     quietly `m' `y' `X1' , aux(`X2')
3.     estimates store `m'3
4. }

. use Data_MPP_small, clear
. quietly replace law=law/100
. foreach m of local method {
2.     quietly `m' `y' `X1' , aux(`X2')
3.     estimates store `m'4
4. }

. estimates table bma3 bma4 walsns3 walsns4 wals3 wals4, ///
>     b(%7.4f) se(%7.4f)
```

Variable	bma3	bma4	walsns3	walsns4	wals3	wals4
_cons	0.0492	0.0492	0.0594	0.0577	0.0618	0.0618
	0.0229	0.0229	0.0221	0.0221	0.0218	0.0218
gdp60	-1.3865	-0.0139	-1.5630	-0.0152	-1.5650	-0.0157
	0.3498	0.0035	0.3270	0.0033	0.3144	0.0031
equipinv	0.1644	0.1644	0.1555	0.1605	0.1582	0.1582
	0.0615	0.0615	0.0551	0.0550	0.0544	0.0544
school60	0.0160	0.0160	0.0175	0.0173	0.0167	0.0167
	0.0102	0.0102	0.0097	0.0097	0.0097	0.0097
life60	0.0008	0.0008	0.0009	0.0009	0.0009	0.0009
	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
dpop	0.1654	0.1654	0.2651	0.2452	0.2714	0.2714
	0.2770	0.2770	0.2487	0.2485	0.2425	0.2425
law	0.0109	1.0890	0.0147	1.2971	0.0134	1.3411
	0.0093	0.9329	0.0065	0.6405	0.0058	0.5804
tropics	-0.0035	-0.0035	-0.0055	-0.0056	-0.0060	-0.0060
	0.0047	0.0047	0.0037	0.0038	0.0035	0.0035
ave1f	-0.0021	-0.0021	-0.0053	-0.0047	-0.0077	-0.0077
	0.0047	0.0047	0.0048	0.0051	0.0051	0.0051
confuc	0.0612	0.0612	0.0443	0.0472	0.0465	0.0465
	0.0185	0.0185	0.0163	0.0164	0.0143	0.0143

legend: b/se

These results show that estimates from `bma` and `wals` are equivariant to scale transformations of focus and auxiliary regressors. On the other hand, the original WALS estimates (i.e. the estimates from the fictitious command `walsns`) are equivariant to scale transformations of the focus regressors, but not to scale transformations of the auxiliary regressors. Obviously, similar considerations hold for the WALS estimator based on Subbotin prior. As an example, we show the two variants of the WALS estimates using a neutral Subbotin prior with  $q = 0.5$ .

```
. use Data_MPP_small, clear
. walsns `y' `X1' , aux(`X2') q(.5)
WALS estimates - Subbotin prior          Number of obs =    74
k1                                       =          6
k2                                       =          4
q                                       =    0.5000
c                                       =    1.6783
```

growth	Coef.	Std. Err.	t	[2-Std. Err. Bands]
			kappa	= 4.3
_cons	.0585353	.0222	2.64	.0363352 .0807353
gdp60	-.0155704	.0033093	-4.70	-.0188797 -.012261
equipinv	.1497584	.0556538	2.69	.0941045 .2054122
school60	.01755	.0096988	1.81	.0078512 .0272488
life60	.0008474	.0003515	2.41	.0004959 .0011989
dpop	.2777177	.2523091	1.10	.0254086 .5300268
law	.0158902	.0068449	2.32	.0090454 .0227351
tropics	-.0055956	.0039901	-1.40	-.0095856 -.0016055
avelf	-.0049763	.0046098	-1.08	-.0095861 -.0003665
confuc	.0467552	.0174805	2.67	.0292747 .0642357

```
. wals `y' `X1' , aux(`X2') q(.5)
WALS estimates - Subbotin prior          Number of obs =    74
                                         k1              =     6
                                         k2              =     4
                                         q                =  0.5000
                                         c                =  1.6783
                                         kappa             =   1.3
```

growth	Coef.	Std. Err.	t	[2-Std. Err. Bands]
_cons	.062339	.0217092	2.87	.0406298 .0840482
gdp60	-.0156395	.0031042	-5.04	-.0187437 -.0125353
equipinv	.1546535	.054792	2.82	.0998615 .2094456
school60	.0162502	.0096439	1.69	.0066064 .0258941
life60	.0008451	.0003503	2.41	.0004947 .0011954
dpop	.2825158	.2411789	1.17	.0413369 .5236948
law	.0138898	.0058776	2.36	.0080122 .0197673
tropics	-.0060841	.0034191	-1.78	-.0095031 -.002665
avelf	-.0087563	.0049474	-1.77	-.0137037 -.0038088
confuc	.0494225	.0140466	3.52	.0353759 .0634692

The estimates from the `walsns` command coincide exactly with those reported in Table 3 of Einmahl et al. (2011) under WALS ( $q = 0.5$ ). The estimates from our `wals` command have the advantage of being scale-equivariant. An additional advantage of our command for WALS with a neutral Subbotin prior is that it allows specifying any real value of the free parameter  $q$  in the interval  $(0, 1)$ . Furthermore, for values of  $q \neq 1$  and  $q \neq 0.5$ , one can also control accuracy the numerical processes required to compute the constrained parameter  $c$  of a Subbotin prior under neutrality.

## 8 BMA with many auxiliary regressors

As discussed in the previous sections, the computational burden of an exact BMA estimator increases exponentially with the number of auxiliary regressors. This section provides some additional insights on this topic by focusing on two issues. First, we would like to assess whether our `bma` command can only support a limited number of auxiliary regressors. Notice that, when  $k_2$  is large,

the most binding constraint is expected to be computing time. BMA estimates are indeed obtained by partial sum over the entire model space but without computing matrices or vectors of dimensions  $2^{k_2}$ . Nevertheless, we cannot exclude a priori that for some large value of  $k_2$  our `bma` command also suffers from out-of-memory problems and numerical errors in the computation of the model weights  $\lambda_i$ . Accordingly, we want to test our `bma` command for a moderately large value of  $k_2$ .

Given that computing time is expected to be a crucial element to establish what is computationally feasible, the second purpose of our analysis is to provide an ex-ante evaluation of the effective time needed for exact BMA estimation of a model with a certain number of auxiliary regressors. For estimating a model with  $n$  observations,  $k_1$  focus regressors and  $k_2$  auxiliary regressors, we suggest the approximation

$$t(k_2) = 2^{k_2} t_0, \quad t_0 = \exp(\tau_0 + \tau_1 k_2 + \tau_2 k_2^2). \quad (16)$$

The computing time  $t$  depends on the number of auxiliary regressors  $k_2$ , conditional on  $n$ ,  $k_1$ , and of course the type of computer. The term  $2^{k_2}$  is the dimension of the model space, and the term  $t_0$  represents an average measure of the computing time needed for estimating a single model. The latter is expressed as a quadratic function of  $k_2$  to capture the effects of operations that are independently, linearly and quadratically related to the number of auxiliary regressors. The parameters  $\tau_j$ ,  $j = 0, 1, 2$ , can be easily estimated by non-linear least squares using information on the effective computing time for a range of feasible values of  $k_2$ . These estimates can then be used to predict the computing time needed for estimation of a model with the desired number of auxiliary regressors.

To shed some light on these two topics, we use the same data set analyzed by Sala-I-Martin et al. (2004), Ley and Steel (2007) and Magnus et al. (2010) which includes 67 determinants of the average GDP growth per-capita between 1960 and 1996 for 88 countries.<sup>7</sup> In the spirit of the BMA approach advocated by Magnus et al. (2010), we treat 7 of the 67 growth determinants as focus regressors and the remaining as auxiliary regressors. The dimension of the underlying model space is  $I = 2^{60} = 1.15 \times 10^{18}$ . Even if we assume that each model could be estimated in  $1 \times 10^{-9}$  seconds, exact BMA estimation over all possible models would require more than 1000 years. We must necessarily consider a smaller subset of auxiliary variables. In order to select the auxiliary regressors which are more robustly correlated with growth, we first ordered these variables by the WALS estimates of their  $t$ -ratios in absolute value. Then, we carried out exact BMA estimation

---

<sup>7</sup> Data can be downloaded from the website [www.janmagnus.nl/items/BMA.pdf](http://www.janmagnus.nl/items/BMA.pdf). For a description of the data see Sala-I-Martin et al. (2004) and Magnus et al. (2010).

with  $k_2$  ranging from 10 to 20 for estimating the parameters  $\tau_j$  and so the maximum number of auxiliary regressors allowed in a certain amount of time. Using a desktop computer with two quad-core Intel Xeon E5504/2 GHz processors and Stata MP4 version 11.2, we obtained  $\hat{\tau}_0 = -17.19$ ,  $\hat{\tau}_1 = .06$  and  $\hat{\tau}_2 = -.35 \times 10^{-3}$ . On the basis of these estimates, we decided to set the maximum value of  $k_2$  at 30 with an expected computing time of 153 hours (i.e. 6 days and 9 hours).

Predicted and effective computing time for  $k_2$  ranging from 20 to 30 are plotted in Figure 2. We can see that the proposed approximation allows predicting the effective computing time accurately. The time needed for estimating the model with  $k_2 = 30$  was 157 hours (i.e. 6 days and 13 hours). BMA and WALS estimates of the focus parameters for the specifications with  $k_2$  equal to 20, 25 and 30 are presented in Table 1. For WALS, we also provide estimates of the specification with  $k_2 = 60$  and estimates based on different prior distributions (Laplace and Subbotin with  $q = 0.5$ ). A number of interesting findings are worth noticing. First, our `bma` command allows performing exact BMA estimation with a moderately large set of auxiliary regressors (at least  $k_2 = 30$ ). We do not exclude that our command works properly with  $k_2 > 30$ , but this would require either a faster computer or a considerably larger amount of computing time. Second, BMA and WALS estimates can be subject to non-negligible differences. For example, in the specification with  $k_2 = 30$ , we find that the estimate and the standard error of the constant term in WALS are two times larger than those obtained in BMA. On the other side, differences between WALS estimates based on Laplace and Subbotin priors appear to be negligible. Third, the precision of these model averaging estimators decreases with the number of auxiliary variables because of both the greater model uncertainty and the higher degree of collinearity among explanatory variables. A comparison of the WALS estimates for the model with  $k_2 = 30$  and  $k_2 = 60$  also suggests that selecting smaller subsets of auxiliary regressors may lead to severely understated standard errors.

## 9 BMA with many observations

So far, we considered two empirical applications on GDP growth which typically involve a relatively small sample size. In this section, we investigate performances of our `bma` command for empirical applications involving a considerably larger sample size. When the sample size is large, the first important improvement of our `bma` command is related to the normalization of the model weights. In order to emphasize this issue, we consider a simulated experiment involving two designs with different sample size:  $n = 100$  in the first design and  $n = 1000$  in the second design. The true data generating process for the outcome variable is always an intermediate model between the restricted

and the unrestricted models.

```

. local sampsize 100 1000
. foreach n of local sampsize {
2.   clear all
3.   quietly set obs `n'
4.   set seed 123
5.   drawnorm x1 x2_1 x2_2 x2_3 x2_4 x2_5          ///
>     x2_6 x2_7 x2_8 x2_9 eps, n(`n')
6.   gen y = 1 + x1 + x2_1 + x2_2 + x2_3 + x2_4 + x2_5  ///
>     + x2_6 + x2_7 + eps
7.   bma y x1, aux(x2_*) nodots
8. }

```

BMA estimates

						Number of obs =	100
						k1	=
						k2	=
							9
y	Coef.	Std. Err.	t	pip	[2-Std. Err. Bands]		
_cons	.7781606	.1057033	7.36	1.00	.6724573	.8838639	
x1	1.183032	.1115203	10.61	1.00	1.071512	1.294552	
x2_1	.9805816	.1099803	8.92	1.00	.8706013	1.090562	
x2_2	.9353698	.1070002	8.74	1.00	.8283696	1.04237	
x2_3	1.055535	.1139422	9.26	1.00	.9415929	1.169477	
x2_4	.9668652	.1105861	8.74	1.00	.8562791	1.077451	
x2_5	.9973639	.0973275	10.25	1.00	.9000364	1.094691	
x2_6	.9370233	.1106242	8.47	1.00	.8263992	1.047648	
x2_7	.9637105	.0925731	10.41	1.00	.8711374	1.056284	
x2_8	.0032232	.0323011	0.10	0.10	-.0290779	.0355243	
x2_9	.0662094	.1046876	0.63	0.37	-.0384782	.170897	

BMA estimates

						Number of obs =	1000
						k1	=
						k2	=
							9
y	Coef.	Std. Err.	t	pip	[2-Std. Err. Bands]		
_cons	.9929054	.0318054	31.22	1.00	.9611001	1.024711	
x1	.9600091	.031499	30.48	1.00	.9285101	.9915081	
x2_1	.9592167	.0331751	28.91	1.00	.9260416	.9923918	
x2_2	.9642845	.0301792	31.95	1.00	.9341053	.9944636	
x2_3	1.009436	.03134	32.21	1.00	.9780958	1.040776	
x2_4	.9992829	.0327001	30.56	1.00	.9665828	1.031983	
x2_5	.9458756	.0328291	28.81	1.00	.9130465	.9787048	
x2_6	.9940521	.0320843	30.98	1.00	.9619678	1.026136	
x2_7	1.01109	.0327991	30.83	1.00	.9782904	1.043889	
x2_8	.00065	.0067813	0.10	0.04	-.0061313	.0074314	
x2_9	-.0002898	.0060394	-0.05	0.03	-.0063292	.0057496	

Our simulated data consist of 10 explanatory variables and a random error independently drawn from standardized Gaussian distributions. The true model for the outcome variable includes a constant term and only 8 of the 10 explanatory variables available in the data. All regression parameters are set to 1. BMA estimation is carried out by treating the constant term and x1 as focus regressors and x2\_1–x2\_9 as auxiliary regressors. Our bma command is not affected by numerical problems and provides satisfactory estimates in both designs. If we try to estimate the same model with the original Matlab command for BMA estimation, then we obtain the same



estimates when  $n = 100$  but unfeasible estimates when  $n = 1000$ . In the second design, the residual sum of squares from the restricted model is numerically too large and so the  $\lambda_i^*$  explode.

To show the other computational advantages of our `bma` command in cases where the sample size is large, we consider the empirical application of Dardanoni et al (2011b) who apply BMA and WALS in the context of a linear regression model where some covariate values are missing but imputations are available to fill-in the missing values.<sup>8</sup> In this context, the availability of imputations generates a trade-off between bias and precision: the complete cases are often too few, so precision is lost, but filling-in the missing values with the imputations may lead to bias. Dardanoni et al. (2011b) show that this bias-precision trade-off is equivalent to that arising in an extended regression model with two subsets of regressors: the focus regressors corresponding to the observed and imputed covariates, and the auxiliary regressors corresponding to all possible interactions between the focus regressors and a set of indicators for the missing-data patterns. Their empirical application focuses on a linear regression model for the body mass index (BMI) of 50+ European men using a sample of 11475 observations from the Survey of Health, Ageing, and Retirement in Europe (SHARE).<sup>9</sup> The model includes 6 focus regressors, of which 4 are fully observed (the constant term, age, age squared and a dummy for not having a high school degree) and 2 are imputed (household income and food expenditure). In addition to the subsample with complete data, there are 3 missing data patterns and so 18 auxiliary regressors. Our BMA estimates (not presented here) coincide exactly with those obtained by Dardanoni et al. (2011b) using the original Matlab command for BMA estimation. The WALS estimates are slightly different because of the preliminary scaling step introduced by our `wals` command. We also notice that, in this application,  $\kappa$  decreases from 202.3 to 23.3. Thus, our WALS estimates are also more accurate than those obtained with the original Matlab command.

Finally, we want to investigate the relationship between computing time and sample size in BMA estimation. Accordingly, we randomly drew from the original data 10 subsamples of sizes ranging from a minimum of  $n = 500$  to a maximum of  $n = 5000$ . For each subsample, we computed BMA estimates in Stata and Matlab using the same desktop computer.<sup>10</sup> The effective computing time required by the Stata and the Matlab commands for BMA estimation are plotted in the left

---

<sup>8</sup> For a Stata implementation of this approach see the `gmi` command of Dardanoni et al (2011a).

<sup>9</sup> Data can be downloaded from the SHARE Research Data Center: <http://www.share-project.org/>. To get access to the data, researchers have to complete a statement concerning the use of the microdata.

<sup>10</sup> This exercise was performed on a desktop computer with one dual-core Intel GX620/3.4 GHz processor. The operating system is Microsoft Windows XP Home edition. For Stata, we used version 11.2 - MP2. For Matlab, we used version 7.8.0.

subpanel of Figure 3. The right subpanel of the same figure shows the relative performances of Matlab *versus* Stata. We can see that the effective computing time required by the original Matlab command increases quadratically with the sample size, while the effective computing time required by our Stata command increases linearly. For the subsample with  $n = 5000$ , our `bma` command is about 35 times faster than the original Matlab command. Because of out-of-memory problems, we cannot obtain the Matlab estimates of this model on the entire sample with  $n = 11475$ . The Stata estimates are obtained in 2 hours.

## 10 Conclusions

In this article, we have introduced the new Stata commands `bma` and `wals` which implement the BMA and WALS estimators developed by Magnus et al. (2010). Unlike standard pretest estimators, these model averaging techniques allow estimating linear regression models with uncertainty about the choice of the explanatory variables by taking into account both the model selection and the estimation steps. Although the `bma` and `wals` commands are written on the basis of the original Matlab commands, the BMA and WALS algorithms have been improved in several respects. The `bma` command is faster than the corresponding Matlab command, especially when the sample size is large, and it uses a more stable normalization of the model weights. The `wals` command is scale-equivariant, is more accurate than the corresponding Matlab command, and allows using more flexible specifications of the prior distributions. The empirical applications considered in the article suggest that performances of the our Stata commands are superior to those of the original Matlab commands.

## References

- Cheney, W., and D. Kincaid, ed. 2008. *Numerical Mathematics and Computing*, 6th ed. Belmont, USA: Thomson Brooks/Cole.
- Danilov, D. 2005. (2006), Estimation of the Mean of a Univariate Normal Distribution when the Variance Is not Known. *Econometrics Journal* 8: 277–291.
- Danilov, D., and J.R. Magnus. 2004. On the Harm that Ignoring Pretesting Can Cause. *Journal of Econometrics* 122:27–46.
- Dardanoni, V., G. De Luca, S. Modica, and F. Peracchi. 2011a. A Generalized Missing Indicator Approach to Regression with Imputed Covariates. Technical report, EIEF.
- Dardanoni, V., S. Modica, and F. Peracchi. 2011b. Regression with Imputed Covariates: A Generalized Missing Indicator Approach. *Journal of Econometrics* doi: 10.1016/j.jeconom.2011.02.005.
- Einmahl, J.H.J., K. Kumar, and J.R. Magnus. 2011. On the Choice of Prior in Bayesian Model Averaging. Technical report, Tilburg University.
- Fernández, C., E. Ley, and M.F.J. Steel. 2001. Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics* 16: 563–576.
- Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Vollinsky. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science* 14: 382–417.
- Leamer, E.E., ed. 1978. *Specification Search: Ad Hoc Inference with Nonexperimental Data*. New York, USA: Wiley.
- Lehmann, E.L., and G. Casella, ed 1998. *Theory of Point Estimation*. 2nd ed. New York, USA: Springer.
- Ley, E., and M.F.J. Steel. 2007. “Jointness in Bayesian Variable Selection with Applications to Growth Regression”. *Journal of Macroeconomics* 29: 476–493.
- Magnus, J.R.. 2002. Estimation of the Mean of a Univariate Normal Distribution with Known Variance. *Econometrica* 67: 639–643.
- Magnus, J.R., and J. Durbin. 1999. Estimation of Regression Coefficients of Interest when Other Regression Coefficients Are of no Interest. *Econometrics Journal* 5: 225–236.
- Magnus, J.R., and H. Neudecker, ed 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, USA: Wiley.
- Magnus, J.R., O. Powell, and P. Prüfer. 2010. A Comparison of Two Model Averaging Techniques with an Application to Growth Empirics. *Journal of Econometrics*, 154: 139–153.
- Magnus J.R., A.T.K. Wan, and X. Zhang. 2011. Weighted Average Least Squares Estimation with Non-spherical Disturbances and an Application to the Hong Kong Housing Market. *Computational Statistics & Data Analysis*, 55: 1331–1341.
- Masanjala, W., and C. Papageorgiou. 2008. Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging. *Journal of Applied Econometrics* 23: 671–682.
- Raftery, A.E. 1995. Bayesian Model Selection in Social Research. *Sociological Methodology* 25: 111–163.
- Sala-i-Martin, X., G. Doppelhofer, and R.I. Miller. 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94: 813–835.
- Zellner, A. 1986. On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, ed. P. K. Goel and A. Zellner, 133–243. Amsterdam: North-Holland.

Table 1: BMA and WALs estimates (and standard errors in parentheses) of focus coefficients using increasing numbers of auxiliary regressors.

Method	Variable	Number of auxiliary variables			
		$k_2 = 20$	$k_2 = 25$	$k_2 = 30$	$k_2 = 60$
BMA	constant	0.053 (0.017)	0.056 (0.019)	0.059 (0.019)	
	p60	0.021 (0.007)	0.020 (0.007)	0.021 (0.007)	
	iprice1	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	
	gdpch60l	-0.010 (0.003)	-0.010 (0.003)	-0.010 (0.003)	
	tropical	-0.009 (0.004)	-0.009 (0.004)	-0.008 (0.005)	
	life060	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	
	confuc	0.039 (0.020)	0.040 (0.021)	0.041 (0.021)	
	avelf	-0.005 (0.006)	-0.005 (0.006)	-0.006 (0.006)	
WALS (Laplace)	constant	0.069 (0.029)	0.095 (0.036)	0.108 (0.038)	0.125 (0.070)
	p60	0.022 (0.007)	0.020 (0.007)	0.022 (0.007)	0.027 (0.017)
	iprice1	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
	gdpch60l	-0.009 (0.003)	-0.009 (0.003)	-0.008 (0.003)	-0.010 (0.007)
	tropical	-0.011 (0.005)	-0.010 (0.006)	-0.009 (0.006)	-0.012 (0.015)
	life060	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)
	confuc	0.050 (0.016)	0.055 (0.016)	0.052 (0.017)	0.049 (0.038)
	avelf	-0.000 (0.006)	0.003 (0.006)	0.001 (0.006)	-0.002 (0.012)
WALS (Subbotin)	constant	0.068 (0.027)	0.093 (0.034)	0.103 (0.037)	0.118 (0.065)
	p60	0.023 (0.007)	0.021 (0.007)	0.022 (0.007)	0.026 (0.016)
	iprice1	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
	gdpch60l	-0.008 (0.003)	-0.009 (0.003)	-0.008 (0.003)	-0.010 (0.007)
	tropical	-0.011 (0.005)	-0.011 (0.006)	-0.010 (0.006)	-0.011 (0.014)
	life060	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)
	confuc	0.048 (0.016)	0.052 (0.016)	0.051 (0.017)	0.048 (0.036)
	avelf	-0.001 (0.005)	0.002 (0.006)	0.001 (0.006)	-0.003 (0.011)

Figure 1: Gaussian, Laplace and Subbotin ( $q = 0.5$ ) densities.

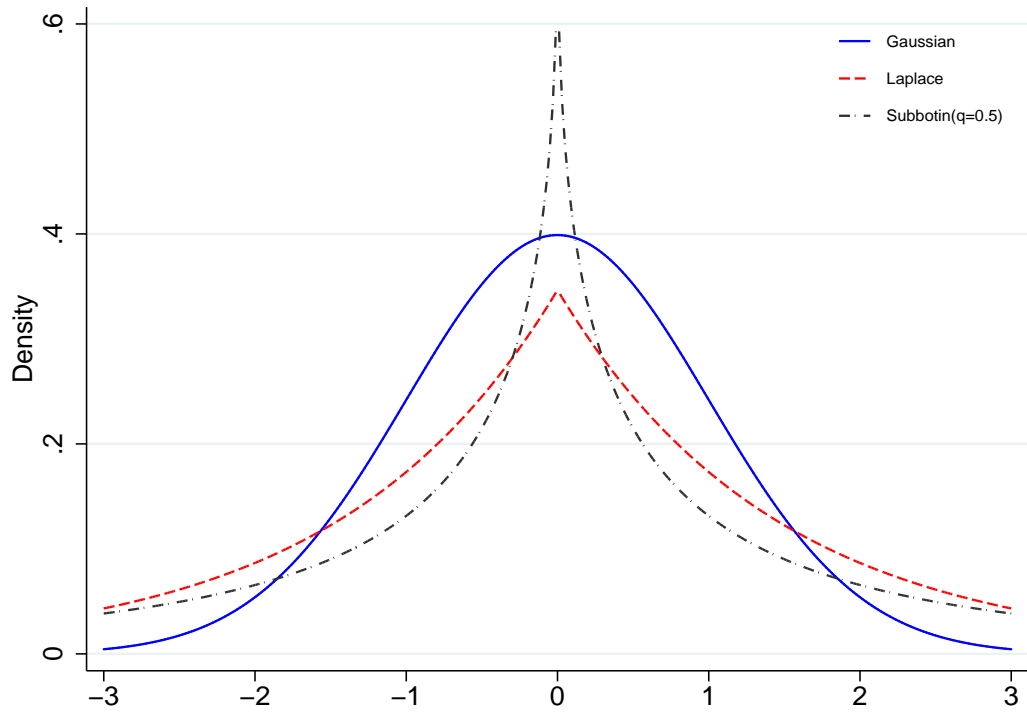


Figure 2: Effective and predicted computing time of the `bma` command as function of the number of auxiliary variables. The dots denote the effective computing time, while the dash-dot line and the shaded area denote the predicted computing time with 95 percent symmetric confidence bands. The sample size is  $n = 88$ .

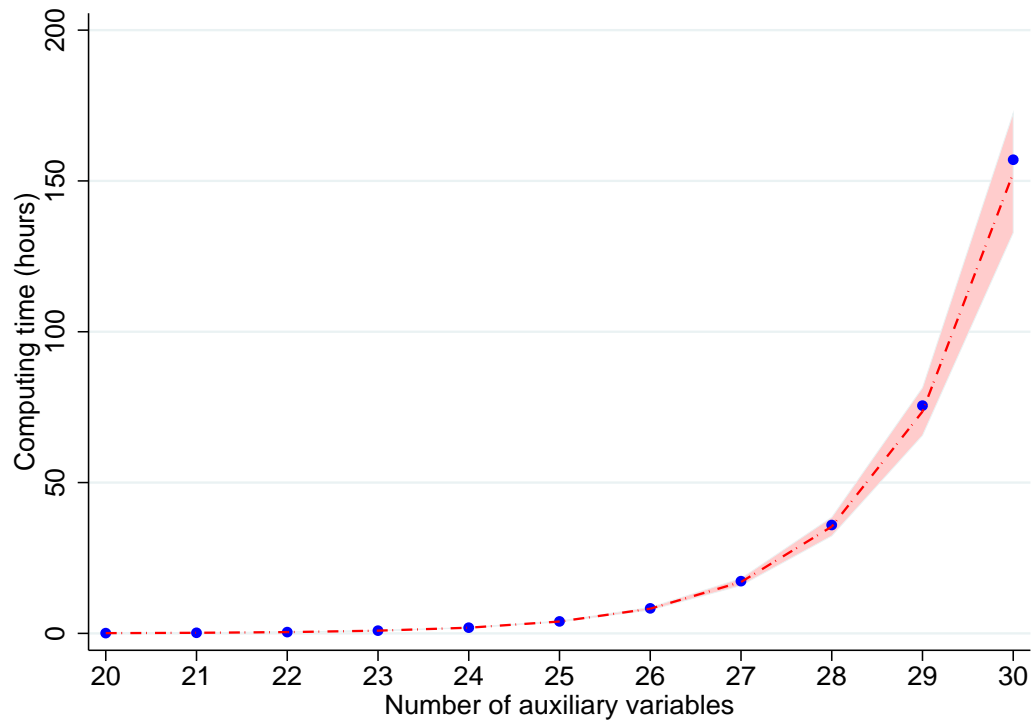


Figure 3: Effective and relative computing time of the Stata and Matlab commands for BMA estimation as function of the number of observations. In each subsample, the number of auxiliary regressors is  $k_2 = 18$ .

