

Bayesian model averaging and weighted-average least squares: Equivariance, stability, and numerical issues

Giuseppe De Luca
ISFOL
Rome, Italy
g.deluca@isfol.it

Jan R. Magnus
Tilburg University
Tilburg, The Netherlands
magnus@uvt.nl

Abstract. In this article, we describe 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 estimator and the weighted-average least-squares estimator developed by [Magnus, Powell, and Prüfer \(2010, *Journal of Econometrics* 154: 139–153\)](#). Unlike standard pretest estimators that 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 several 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: st0239, `bma`, `wals`, model uncertainty, model averaging, Bayesian analysis, exact Bayesian model averaging, weighted-average least squares

1 Introduction

Economic theory provides, in general, some information about the empirical model specification, but it offers little guidance about 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, 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 pretest estimators (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 nonnegligible effects on the statistical properties of our estimators.

In this article, we describe 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 [MSE]). 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, chap. 4, sec. 4–6) and the weighted-average least-squares (WALS) estimator developed by Magnus, Powell, and Prüfer (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 about 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, Powell, and Prüfer (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.¹

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. Section 6 provides some additional remarks about 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.

1. Updated versions of the MATLAB commands for BMA and WALS can be downloaded free of charge from the website <http://www.janmagnus.nl/items/BMA.pdf>.

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 $\mathbf{u} \sim N(0, \sigma^2)$, an $n \times 1$ random vector of unobservable disturbances whose elements are independent and identically distributed. We assume that $k_1 \geq 1$, $k_2 \geq 0$, $k = k_1 + k_2 \leq n - 1$, and 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 that we want in the model because of theoretical reasons or other considerations about the phenomenon under investigation, whereas \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 $\boldsymbol{\beta}_1$, whereas $\boldsymbol{\beta}_2$ is treated as a vector of nuisance parameters. By the properties of partitioned inverses, the unrestricted ordinary least-squares (OLS) estimators of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{1u} &= \hat{\boldsymbol{\beta}}_{1r} - \mathbf{Q}\hat{\boldsymbol{\beta}}_{2u} \\ \hat{\boldsymbol{\beta}}_{2u} &= (\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{M}_1 \mathbf{y} \end{aligned}$$

where $\hat{\boldsymbol{\beta}}_{1r} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{y}$ is the restricted OLS estimator from a regression of \mathbf{y} on \mathbf{X}_1 (with $\boldsymbol{\beta}_2$ 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 MSE sense, the unrestricted OLS estimator $\hat{\boldsymbol{\beta}}_{1u}$ of $\boldsymbol{\beta}_1$. It is a basic result from least-squares theory that by restricting some elements of $\boldsymbol{\beta}_2$ to zero, we can obtain an estimator of $\boldsymbol{\beta}_1$, which is subject to omitted variable bias but is also more precise than the unrestricted OLS estimator $\hat{\boldsymbol{\beta}}_{1u}$. 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.

Because 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 the i th model in the model space by \mathcal{M}_i , which is obtained by including only a subset of k_{2i} auxiliary regressors (with $0 \leq k_{2i} \leq k_2$). Model \mathcal{M}_i is represented as

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_{2i}\boldsymbol{\beta}_{2i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, I \quad (2)$$

where \mathbf{X}_{2i} is an $n \times k_{2i}$ matrix of observations on the included subset of k_{2i} auxiliary regressors, $\boldsymbol{\beta}_{2i}$ is the corresponding subvector of auxiliary parameters, and $\boldsymbol{\epsilon}_i$ 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 β_1 is given by

$$\widehat{\beta}_1 = \sum_{i=1}^I \lambda_i \widehat{\beta}_{1i}$$

where the λ_i are nonnegative random weights that add up to one and $\widehat{\beta}_{1i}$ is the estimate of β_1 obtained by conditioning on model \mathcal{M}_i . In model averaging, the weights λ_i are random because they reflect our confidence in model \mathcal{M}_i based on prior beliefs and data-based diagnostic tests. The choice of these weights depends on the method of estimation. Below we discuss two model-averaging estimators.

3.1 BMA

The BMA estimator developed in Magnus, Powell, and Prüfer (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 (2) can be written as

$$p(\mathbf{y} \mid \beta_1, \beta_{2i}, \sigma^2, \mathcal{M}_i) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{\boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}_i}{2\sigma^2}\right) \quad (3)$$

Prior beliefs on the regression parameters of model \mathcal{M}_i are introduced by imposing conventional noninformative priors on the focus parameters β_1 and the error variance σ^2 , in addition to an informative Gaussian prior on the auxiliary parameters β_{2i} , specifically $\beta_{2i} \mid \beta_1, \sigma^2, \mathcal{M}_i \sim N(0, \sigma^2 V_{0i})$. This leads to a conditional joint prior distribution of the form

$$p(\beta_1, \beta_2, \sigma^2 \mid \mathcal{M}_i) \propto (\sigma^2)^{(k_{2i}+2)/2} \exp\left(-\frac{\beta_{2i}^\top \mathbf{V}_{0i}^{-1} \beta_{2i}}{2\sigma^2}\right) \quad (4)$$

where V_{0i} takes the standard form proposed by Zellner (1986) and Fernández, Ley, and Steel (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 (3) with the conditional joint prior distribution (4) to obtain the conditional posterior distribution $p(\beta_1, \beta_{2i}, \sigma^2 \mid \mathbf{y}, \mathcal{M}_i)$. As argued by Magnus, Powell, and Prüfer (2010), this task

is complicated because 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 (4) 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, Powell, and Prüfer (2010) show that the conditional estimates of β_1 and β_{2i} under model \mathcal{M}_i are given by

$$\begin{aligned}\widehat{\beta}_{1i} &= E(\beta_1 | \mathbf{y}, \mathcal{M}_i) = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{X}_{2i} \widehat{\beta}_{2i}) \\ \widehat{\beta}_{2i} &= E(\beta_{2i} | \mathbf{y}, \mathcal{M}_i) = (1 + g)^{-1} (\mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i})^{-1} \mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{y}\end{aligned}$$

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

$$\begin{aligned}\widehat{\mathbf{V}}_{1i} &= \text{Var}(\beta_1 | \mathbf{y}, \mathcal{M}_i) = s_i^2 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} + \mathbf{Q}_i \widehat{\mathbf{V}}_{2i} \mathbf{Q}_i^\top \\ \widehat{\mathbf{V}}_{2i} &= \text{Var}(\beta_{2i} | \mathbf{y}, \mathcal{M}_i) = s_i^2 (1 + g)^{-1} (\mathbf{X}_{2i}^\top \mathbf{M}_1 \mathbf{X}_{2i})^{-1} \\ \widehat{\mathbf{V}}_{12i} &= \text{Cov}(\beta_1, \beta_{2i} | \mathbf{y}, \mathcal{M}_i) = -\mathbf{Q}_i \widehat{\mathbf{V}}_{2i}\end{aligned}$$

where, under each model \mathcal{M}_i , $s_i^2 = (\mathbf{y}^\top \mathbf{M}_1 \mathbf{A}_i \mathbf{M}_1 \mathbf{y}) / (n - k_1 - 2)$ is the estimate of σ^2 , $\mathbf{Q}_i = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_{2i}$ 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 about the joint prior distribution, one can show that

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

where c is a normalizing constant chosen to guarantee that the λ_i add up to one (see section 6).

Given the conditional estimates $\widehat{\beta}_{1i}$ and $\widehat{\beta}_{2i}$ of the regression parameters of model \mathcal{M}_i and the model weights λ_i , the unconditional BMA estimates of β_1 and β_2 are computed as

$$\hat{\beta}_1 = E(\beta_1 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \hat{\beta}_{1i}$$

$$\hat{\beta}_2 = E(\beta_2 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \mathbf{T}_i \hat{\beta}_{2i}$$

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 $\hat{\beta}_{2i}$ in $k_2 \times 1$ vectors by setting to zero the elements of β_2 , which are excluded from model \mathcal{M}_i . The elements of the posterior variance–covariance matrix are given by

$$\text{Var}(\hat{\beta}_1 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \left(\hat{\mathbf{V}}_{1i} + \hat{\beta}_{1i} \hat{\beta}_{1i}^\top \right) - \hat{\beta}_1 \hat{\beta}_1^\top$$

$$\text{Var}(\hat{\beta}_2 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \mathbf{T}_i \left(\hat{\mathbf{V}}_{2i} + \hat{\beta}_{2i} \hat{\beta}_{2i}^\top \right) \mathbf{T}_i^\top - \hat{\beta}_2 \hat{\beta}_2^\top$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2 | \mathbf{y}) = \sum_{i=1}^I \lambda_i \left(\hat{\mathbf{V}}_{12i} + \hat{\beta}_{1i} \hat{\beta}_{2i}^\top \right) \mathbf{T}_i^\top - \hat{\beta}_1 \hat{\beta}_2^\top$$

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 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 β_2 may not be attractive in situations where no prior information is 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 WALS

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 for 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 $\boldsymbol{\gamma}_2 = \mathbf{\Lambda}^{1/2} \mathbf{P}^\top \boldsymbol{\beta}_2$ such that $\mathbf{Z}_2^\top \mathbf{M}_1 \mathbf{Z}_2 = \mathbf{I}_{K_2}$ and $\mathbf{Z}_2 \boldsymbol{\gamma}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2$. The original vector of auxiliary parameters $\boldsymbol{\beta}_2$ can always be recovered from $\boldsymbol{\beta}_2 = \mathbf{P} \mathbf{\Lambda}^{-1/2} \boldsymbol{\gamma}_2$.

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

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{1u} &= \widehat{\boldsymbol{\beta}}_{1r} - \mathbf{R} \widehat{\boldsymbol{\gamma}}_{2u} \\ \widehat{\boldsymbol{\gamma}}_{2u} &= \mathbf{Z}_2^\top \mathbf{M}_1 \mathbf{y}\end{aligned}$$

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, such that \mathbf{S}_i captures the restrictions placed on $\boldsymbol{\gamma}_2$ under model \mathcal{M}_i , then the restricted OLS estimators of $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_{2i}$ are given by

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{1i} &= \widehat{\boldsymbol{\beta}}_{1r} - \mathbf{R} \mathbf{W}_i \widehat{\boldsymbol{\gamma}}_{2u} \\ \widehat{\boldsymbol{\gamma}}_{2i} &= \mathbf{W}_i \widehat{\boldsymbol{\gamma}}_{2u}\end{aligned}$$

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

The key advantage of these transformations lies in the fact that $\widehat{\boldsymbol{\gamma}}_{2u} \sim N_{k_2}(\boldsymbol{\gamma}_2, \sigma^2 \mathbf{I}_{k_2})$. This result has several implications on the computational aspects and the statistical properties of the WALs estimator. First, under some minimal regularity conditions on the model weights λ_i , the WALs estimator of $\boldsymbol{\beta}_1$ is of the form

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

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

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

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

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

Third, because the k_2 components of γ_2 are independent, they can be estimated separately by exploiting the information that $\widehat{\gamma}_{2j} \sim N(\gamma_{2j}, \sigma^2)$. In Magnus, Powell, and Prüfer (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 from 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.² Furthermore, this Bayesian estimator is based on a Laplace prior distribution,

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

with $c = \log 2$ to satisfy the property of neutrality (that is, the prior median of η_j is zero and the prior median of η_j^2 is one), which reflects our notion of ignorance in situations where we do not know whether the t ratio η_j is greater than or less than one in absolute value.

The WALs estimator proposed by Einmahl, Kumar, and Magnus (2011) uses instead an estimator $\widetilde{\eta}_j$ of η_j based on the Subbotin density

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

with $c > 0$ and $0 < q < 1$. This prior allows for obtaining a class of estimators $\widetilde{\eta}_j(q, c)$ with better properties than the Laplace estimator $\widehat{\eta}_j$, especially when η_j is large. As for the choice of the parameters q and c , Einmahl, Kumar, and Magnus (2011) show that q must belong to the interval $(0, 1)$ to obtain a robust estimator of η_j . Given $q = \bar{q}$, the parameter c can be chosen implicitly by solving the nonlinear (5) to satisfy neutrality.

$$\int_0^1 \pi(\eta; \bar{q}, c) d\eta = \frac{1}{4} \quad (5)$$

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 with a density that is less flat in the interval $(0, 1)$ and has thicker tails. For empirical applications, Einmahl, Kumar, and Magnus (2011) recommend using a Subbotin prior with $q = 0.5$.

2. Notice when estimating η_j , the unknown parameter σ^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.

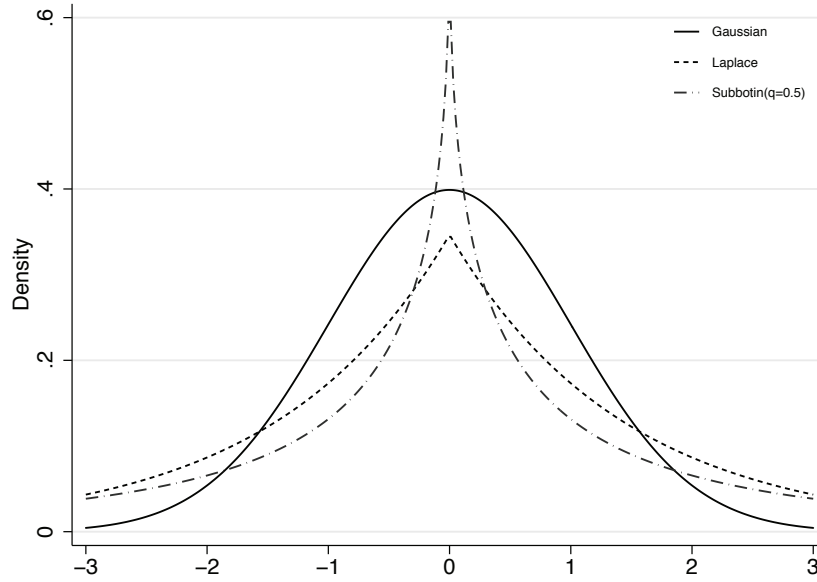


Figure 1. Gaussian, Laplace, and Subbotin ($q = 0.5$) densities

Let us denote by $\bar{\eta}$ the Laplace or the Subbotin estimator of $\eta = (\eta_1, \dots, \eta_{k_2})$. Magnus, Powell, and Prüfer (2010) show that the WALS estimators of the regression parameters β_1 and β_2 are given by

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

and the elements of their variance–covariance matrix are

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

where $\mathbf{Q} = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2$ and $\mathbf{\Omega}$ is the diagonal variance–covariance matrix of $\bar{\eta}$. This model-averaging technique can be generalized to nonspherical errors (see Magnus, Wan, and Zhang [2011]), though this generalization is not included in our current WALS code.

4 Equivariance

An estimator may or may not be equivariant to a certain transformation.³ If the transformation is considered to be “trivial”, then we prefer the estimator to be equivariant, that is, to not 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 nontrivial 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 nontrivial 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 σ^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 true 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 (1999, chap. 14, sec. 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; 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, then neither \mathbf{M}_1 nor $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$ is affected, and therefore BMA and WALS estimates are both equivariant to centering. The reason is simple. Suppose that the first column of \mathbf{X}_1 is \mathbf{z} , 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 nonsingular $k_1 \times k_1$ matrix

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

3. For a formal treatment of the principle of equivariance, see Lehmann and Casella (1998, chap. 3).

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,

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

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

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

because $\mathbf{M}_1 \boldsymbol{\nu} = \mathbf{0}$. 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 kilograms rather than in grams, then we expect nothing to change other than that β_2 is multiplied by 1,000. In a standard (non-Bayesian) context, the OLS estimator is scale-independent, but in a Bayesian context, the OLS estimator is scale-independent 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 semiorthogonalization, which gives us great benefits but makes the estimator scale-dependent because the orthogonal matrix \mathbf{P} and the diagonal matrix $\boldsymbol{\Lambda}$ will depend on the scaling in a nontrivial (nonlinear) 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. This also stabilizes both matrices such that inversion and eigenvalue routines are numerically more stable. The only effect of the scaling in \mathbf{X}_1 is numerical stability, but the scaling in \mathbf{X}_2 has two effects: numerical stability and scale-independence. The property of scale-independence is particularly important to ensure that the WALS estimates of the focus and auxiliary parameters are interpretable. Without our preliminary step, these estimates would depend on scaling of the auxiliary variables in a nonlinear way (see the examples in section 7).

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]
```

depvar is the dependent variable and *varlist* is the optional list of focus regressors (including the constant term, if any) that are included with certainty in the model. Both commands are programmed in Mata on the basis of the original MATLAB commands associated with Magnus, Powell, and Prüfer (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`

`auxiliary`(*varlist*) is the required list of auxiliary regressors of which we are less certain.

`nodots` suppresses the display of the dots that track the progress of `bma` estimation.

Dots are displayed only if the model space consists of more than 128 models (that is, at least seven auxiliary regressors). One dot means that 1% of the models in the model space has been fit.

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

`noconstant` specifies that the constant term be 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`

`auxiliary`(*varlist*) is the required list of auxiliary regressors of which we are less certain.

`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. 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 nonlinear equation for the constrained parameter c of a Subbotin density under neutrality. The default uses 10,000 data points. When $q = 1$ or $q = 0.5$, this option is ineffective because the solution of the constrained parameter c is determined analytically.

`eps`(#) specifies the convergence criterion used by the built-in Stata command `n1` when solving the nonlinear equation for the constrained parameter c of a Subbotin density under neutrality. The default is `eps(1e-8)`. When $q = 1$ or $q = 0.5$, this option is ineffective because the solution of the constrained parameter c is determined analytically.

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

$q = 1$ or $q = 0.5$, this option is ineffective because the solution of the constrained parameter c is determined analytically.

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

6 Additional remarks

1. `bma` improves the original MATLAB command for BMA estimation in two ways. First, `bma` 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

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

If the sample size is large, this normalization may lead to model weights that are numerically too large 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 `bma`, 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 λ_i^* are always bounded in the $(0, 1)$ interval. Second, whenever the sample size is moderately large (for example, $n > 100$), `bma` is considerably faster and requires much less memory than the corresponding MATLAB command because it avoids computing $n \times n$ matrices.

2. `wals` improves the original MATLAB command for WALS estimation in two ways. 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 for specifying neutral Subbotin priors with any real value of q in the interval $(0, 1)$ instead of a list of nine focal values $q = (0.1, 0.2, \dots, 0.9)$. `wals` 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.⁴
3. `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 dataset, and they compute estimated covariances between focus and auxiliary parameters.

4. A description of these alternative quadrature methods can be found in [Cheney and Kincaid \(2008\)](#).

4. `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 the MSE of the restricted OLS estimator (see Magnus [2002]). The intuition behind this result is also related to a well-known property of \bar{R}^2 (the adjusted R^2), which rises if and only if the t ratio associated with an added regressor is greater than one in absolute value. On the basis of this criterion, our commands provide one-standard error bands of the estimated regression parameters.

7 Example

This section uses the growth data analyzed by Magnus, Powell, and Prüfer (2010) to illustrate `bma` and `wals`, validate their estimation results, and investigate equivariance of the BMA and WALS estimators to shift and scale transformations of the explanatory variables.⁵ The data constitute a cross section of the average growth rate of the per capita gross domestic product (GDP) of 74 countries from 1960–1996.

```
. use data_mpp_small
. describe
Contains data from data_mpp_small.dta
obs:           74
vars:          11                    7 Nov 2011 13:15
size:         4,218
```

variable name	storage type	display format	value 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:
```

5. Instructions to download the original data can be found at <http://www.janmagnus.nl/items/BMA.pdf>.

```
. 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, Powell, and Prüfer (2010) provide BMA and WALS estimates of different model specifications to test the implications of alternative growth theories. Here, for simplicity, we focus on setup 1 of their model 1, which allows for 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, and 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

growth	Coef.	Std. Err.	t	pip	[1-Std. Err. Bands]
_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
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 `bma` provides information about estimated coefficients and their standard errors (mean and standard deviation of the posterior distribution), t ratios, pos-

terior inclusion probabilities (the posterior probability that a variable is included in the model), and one-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. Estimated coefficients and standard errors match those reported in table 2 of Magnus, Powell, and Prüfer (2010) under BMA. An auxiliary regressor is considered to be robustly correlated with the outcome if the t ratio on its coefficient is greater than one in absolute value or, equivalently, the corresponding one-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 `wals`, 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	[1-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	[1-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 `wals` is similar to that of `bma`. The main difference is that WALS does not allow for computing the posterior inclusion probabilities because this model-averaging technique considers only k_2 linear combinations of the model weights λ_i . We can see that the estimates from the fictitious `walsns` command match those reported in table 2 of Magnus, Powell, and Prüfer (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 nonlinear 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 a 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 λ_{\max} and λ_{\min} denote, respectively, the maximum and minimum eigenvalues of $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$. The larger is κ , the more ill-conditioned is the matrix $\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2$. In other words, a large value of κ 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 κ decreases from 4.3 to 1.3 after scaling.

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
equipinv	-0.0139	-0.0139	-0.0156	-0.0156	-0.0157	-0.0157
school60	0.0035	0.0035	0.0033	0.0033	0.0031	0.0031
life60	0.1644	0.1644	0.1555	0.1555	0.1582	0.1582
dpop	0.0615	0.0615	0.0551	0.0551	0.0544	0.0544
tropics	0.0160	0.0160	0.0175	0.0175	0.0167	0.0167
avelf	0.0102	0.0102	0.0097	0.0097	0.0097	0.0097
confuc	0.0008	0.0008	0.0009	0.0009	0.0009	0.0009
	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
	0.1654	0.1654	0.2651	0.2651	0.2714	0.2714
	0.2770	0.2770	0.2487	0.2487	0.2425	0.2425
	0.0109	0.0109	0.0147	0.0147	0.0134	0.0134
	0.0093	0.0093	0.0065	0.0065	0.0058	0.0058
	-0.0035	-0.0035	-0.0055	-0.0055	-0.0060	-0.0060
	0.0047	0.0047	0.0037	0.0037	0.0035	0.0035
	-0.0021	-0.0021	-0.0053	-0.0053	-0.0077	-0.0077
	0.0047	0.0047	0.0048	0.0048	0.0051	0.0051
	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, whereas 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 $0.0492403 + 7.525295 \times (-0.0138652) = -0.05509942$. The effects of scale transformations can be similarly assessed. 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
avelf	-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. The original WALS estimates (estimates from the fictitious command `walsns`) are equivariant to scale transformations of the focus regressors but not to scale transformations of the auxiliary regressors. Similar considerations hold for the WALS estimator based on Subbotin prior.

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 about this topic by focusing on two issues. First, we would like to assess whether `bma` can support only a limited number of auxiliary regressors. When k_2 is large, the most binding constraint is expected to be computing time. BMA estimates are 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 , `bma` also suffers from out-of-memory problems and numerical errors in the computation of the model weights λ_i . Accordingly, we want to test `bma` 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 fitting a model with n observations, k_1 focus regressors, and k_2 auxiliary regressors, we suggest the approximation

$$t(k_2) = 2^{k_2} t_0, \quad t_0 = \exp(\tau_0 + \tau_1 k_2 + \tau_2 k_2^2) \quad (6)$$

The computing time t depends on the number of auxiliary regressors k_2 , conditional on n , k_1 , and 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 to fit one 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 τ_j , $j = 0, 1, 2$, can be easily estimated by nonlinear least squares using information about 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 to estimate a model with the desired number of auxiliary regressors.

To shed some light on these two topics, we use the same dataset analyzed by Sala-i-Martin, Doppelhofer, and Miller (2004), Ley and Steel (2007), and Magnus, Powell, and Prüfer (2010), which includes 67 determinants of the average GDP growth rate per capita of 88 countries from 1960–1996.⁶ In the spirit of the BMA approach advocated by Magnus, Powell, and Prüfer (2010), we treat 7 of the 67 growth determinants as focus regressors and the remaining determinants 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 fit in 1×10^{-9} seconds, the exact BMA estimation over all possible models would require more than 1,000 years. We must necessarily consider a smaller subset of auxiliary variables.

To select the auxiliary regressors that 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 with k_2 ranging from 10 to 20 for estimating

6. Instructions to download the original data can be found at <http://www.janmagnus.nl/items/BMA.pdf>.

the parameters τ_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 $\hat{\tau}_0 = -17.19$, $\hat{\tau}_1 = 0.06$, and $\hat{\tau}_2 = -0.35 \times 10^{-3}$. 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 (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 for predicting the effective computing time accurately. The time needed for fitting the model with $k_2 = 30$ was 157 hours (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 $q = 0.5$).

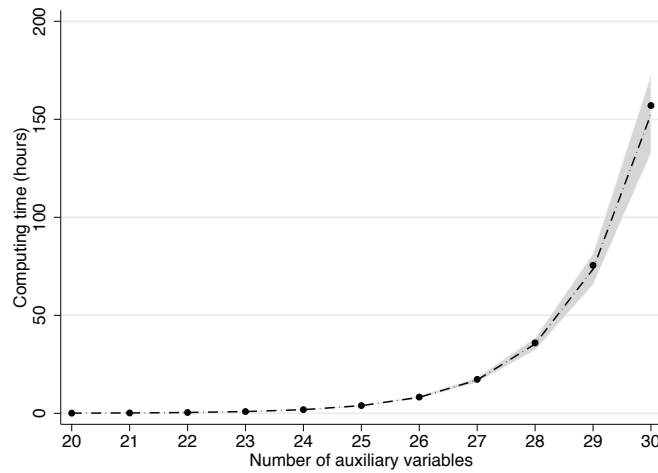


Figure 2. Effective and predicted computing time of `bma` as a function of the number of auxiliary variables. The dots denote the effective computing time. The dash-dot line and the shaded area denote the predicted computing time with 95% symmetric confidence bands. The sample size is $n = 88$.

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)	
	gdpch601	-0.010 (0.003)	-0.010 (0.003)	-0.010 (0.003)	
	tropicalcar	-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)
	gdpch601	-0.009 (0.003)	-0.009 (0.003)	-0.008 (0.003)	-0.010 (0.007)
	tropicalcar	-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)
	gdpch601	-0.008 (0.003)	-0.009 (0.003)	-0.008 (0.003)	-0.010 (0.007)
	tropicalcar	-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)

Several interesting findings are worth noticing. First, **bma** allows for performing exact BMA estimation with a moderately large set of auxiliary regressors (at least $k_2 = 30$). We do not exclude that **bma** works properly with $k_2 > 30$, but this would require either a faster computer or a considerably longer computing time. Second, with the exception of the constant term, the differences between BMA and WALS estimates 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 about GDP growth that involve a relatively small sample size. In this section, we investigate performances of **bma** for empirical applications involving a considerably larger sample size. When the sample size is large,

the first important improvement of `bma` is a more stable normalization of the model weights. To emphasize this issue, we consider a simulated experiment involving two designs with different sample sizes: $n = 100$ 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 sampsiz 100 1000
. foreach n of local sampsiz {
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.     generate 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                =     2
                                                k2                =     9
```

y	Coef.	Std. Err.	t	pip	[1-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                =     2
                                                k2                =     9
```

y	Coef.	Std. Err.	t	pip	[1-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 one. BMA estimation is carried out by treating the constant term and `x1` as focus regressors and `x2_1`–`x2_9` as auxiliary regressors. `bma` is unaffected 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 infeasible estimates when $n = 1000$. In the second design, the residual sum of squares from the restricted model is numerically too large and therefore the λ_i^* explode.

To show the other computational advantages of `bma` in cases where the sample size is large, we consider the empirical application of Dardanoni, Modica, and Peracchi (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.⁷ 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, Modica, and Peracchi (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 of European men aged 50 years and over using a sample of 11,475 observations from the Survey of Health, Ageing, and Retirement in Europe (SHARE).⁸ The model includes six focus regressors, of which four are fully observed (the constant term, age, age squared, and a dummy for not having a high school degree) and two are imputed (household income and food expenditure). In addition to the subsample with complete data, there are three missing-data patterns and therefore 18 auxiliary regressors.

Our BMA estimates (not presented here) match those obtained by Dardanoni, Modica, and Peracchi (2011b) using the original MATLAB command for BMA estimation. The WALS estimates are slightly different because of the preliminary scaling step introduced by `wals`. In this application, we find that κ 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

7. For a Stata implementation of this approach, see the `gmi` command of Dardanoni et al. (2011a).

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

desktop computer.⁹ The relative computing time required by the Stata and the MATLAB commands for BMA estimation are plotted in figure 3. The effective computing time required by the original MATLAB command increases quadratically with the sample size, whereas the effective computing time required by our Stata command increases linearly. The relative computing time of MATLAB versus Stata is then an increasing linear function of the sample size. For the subsample with $n = 5000$, `bma` 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 for the entire sample with $n = 11475$. The Stata estimates are obtained in two hours.

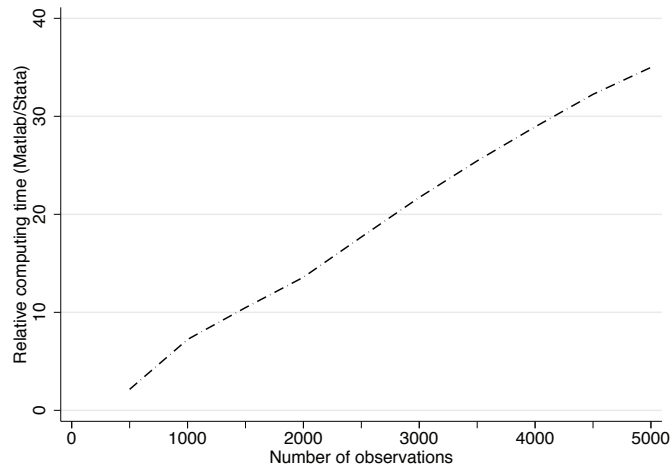


Figure 3. Relative computing time of the Stata and MATLAB commands for BMA estimation as a function of the number of observations. In each subsample, the number of auxiliary regressors is $k_2 = 18$.

10 Conclusions

In this article, we introduced the new Stata commands `bma` and `wals`, which implement the BMA and WALS estimators developed by Magnus, Powell, and Prüfer (2010). Unlike standard pretest estimators, these model-averaging techniques allow for fitting 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 `bma` and `wals` are written on the basis of the original MATLAB commands, the BMA and WALS algorithms have been improved in several ways. `bma` is faster than the corresponding MATLAB command, especially when the sample size is large, and it uses a more stable

9. This exercise was performed using a desktop computer with one dual-core Intel GX620/3.4 GHz processor. The operating system was Microsoft Windows XP Home edition. For Stata, we used Stata/MP2 version 11.2. For MATLAB, we used version 7.8.0.

normalization of the model weights. `wals` is scale-equivariant, is more accurate than the corresponding MATLAB command, and allows for using more flexible specifications of the prior distributions. The empirical applications considered in this article suggest that performance of our Stata commands is superior to those of the original MATLAB commands.

11 Acknowledgments

We thank Valentino Dardanoni, Salvatore Modica, Franco Peracchi, and an anonymous referee for very useful comments. The SHARE data collection has been primarily funded by the European Commission through the 5th, 6th, and 7th framework programs. Additional funding from the U.S. National Institute on Aging as well as from various national sources is gratefully acknowledged (see <http://www.share-project.org> for a full list of funding institutions).

12 References

- Cheney, W., and D. Kincaid. 2008. *Numerical Mathematics and Computing*. 6th ed. Belmont, CA: Brooks/Cole.
- Danilov, D. 2005. 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. <http://www.eief.it/files/2011/05/wp-11-a-generalized-missing-indicator-approach-to-regression-with-imputed-covariates.pdf>.
- Dardanoni, V., S. Modica, and F. Peracchi. 2011b. Regression with imputed covariates: A generalized missing indicator approach. *Journal of Econometrics* 162: 362–368.
- Einmahl, J. H. J., K. Kumar, and J. R. Magnus. 2011. A characterization of Bayesian robustness for a normal location parameter. <http://www.janmagnus.nl/wips/wip23.pdf>.
- 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. Volinsky. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14: 382–417.
- Leamer, E. E. 1978. *Specification Search: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

- Lehmann, E. L., and G. Casella. 1998. *Theory of Point Estimation*. 2nd ed. New York: 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. *Econometrics Journal* 5: 225–236.
- Magnus, J. R., and J. Durbin. 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67: 639–643.
- Magnus, J. R., and H. Neudecker. 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Rev. ed. New York: 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 nonspherical disturbances and an application to the Hong Kong housing market. *Computational Statistics & Data Analysis* 55: 1331–1341.
- Masanjala, W. H., 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, 233–243. Amsterdam: North-Holland.

About the authors

Giuseppe De Luca is a researcher at ISFOL, Rome, Italy.

Jan Magnus is a professor of econometrics at Tilburg University, Tilburg, The Netherlands.