



**HAL**  
open science

## A comparison of two model averaging techniques with an application to growth empirics

Jan R. Magnus, Owen Powell, Patricia Prüfer

► **To cite this version:**

Jan R. Magnus, Owen Powell, Patricia Prüfer. A comparison of two model averaging techniques with an application to growth empirics. *Econometrics*, 2009, 154 (2), pp.139. 10.1016/j.jeconom.2009.07.004 . hal-00613748

**HAL Id: hal-00613748**

**<https://hal.science/hal-00613748>**

Submitted on 6 Aug 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A comparison of two model averaging  
techniques with an application to growth  
empirics

April 2008; this revision: 18 May 2009

Jan R. Magnus\*

*Department of Econometrics & Operations Research,  
Tilburg University, The Netherlands*

Owen Powell

*CentER, Tilburg University, The Netherlands*

Patricia Prüfer

*Netherlands Bureau for Economic Policy Analysis (CPB) and  
CentER, Tilburg University, The Netherlands*

---

\*Corresponding address: Jan R. Magnus, Department of Econometrics & OR, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. *E-mail address:* magnus@uvt.nl.

**Proposed running head:**

Comparison of two model averaging techniques

**Corresponding author:**

Jan R. Magnus

Department of Econometrics and Operations Research

Tilburg University

PO Box 90153

5000 LE Tilburg

The Netherlands

e-mail: magnus@uvt.nl

**Abstract:** Parameter estimation under model uncertainty is a difficult and fundamental issue in econometrics. This paper compares the performance of various model averaging techniques. In particular, it contrasts Bayesian model averaging (BMA) — currently one of the standard methods used in growth empirics — with a new method called weighted-average least squares (WALS). The new method has two major advantages over BMA: its computational burden is trivial and it is based on a transparent definition of prior ignorance. The theory is applied to and sheds new light on growth empirics where a high degree of model uncertainty is typically present.

**JEL Classification:** C51, C52, C13, C11

**Keywords:** Model averaging, Bayesian analysis, Growth determinants

## 1 Introduction

This paper has two purposes. First, it introduces a new model averaging technique, called weighted-average least squares (hereafter WALs), which we claim to be theoretically and practically superior to standard Bayesian model averaging (BMA). It is theoretically superior because it treats our ignorance about the priors in a different manner, thereby obtaining a better risk profile and, in particular, avoiding unbounded risk. It is practically superior because the space over which we need to perform model selection increases *linearly* rather than exponentially in size. Thus, if we have sixty regressors to search over (which is not unusual in the growth literature), then computing time of standard BMA is of the order  $2^{60}$ , while computing time of WALs is of the order 60. This means that what WALs can do in one second, BMA can only do in six hundred million years. Exact computation of a complete BMA is therefore rarely done; instead some Markov chain Monte Carlo (MCMC) method is typically applied.

The second purpose is to contribute to the debate on growth empirics. Since the seminal studies of Kormendi and Meguire (1985) and Barro (1991), empirical research on the determinants of economic growth has identified numerous variables as being robustly (partially) correlated with productivity growth in an economy. Durlauf et al. (2005) list 145 potential right-hand side variables for growth regressions and cluster them into more than forty areas (or theories), such as human capital, finance, government, and trade. Taking into account the limited number of observations available at a national level, growth empirics has been heavily criticized because of the inherent model uncertainty; see Durlauf et al. (2005) for a recent in-depth survey. Sometimes growth theory can support choices of specific variables, but the inclusion or exclusion of most variables is typically arbitrary, a phenomenon labeled the ‘open-endedness’ of growth theory (Brock and Durlauf, 2001). In addition, while theory may provide general qualitative variables (such as human capital), it does not tell us how these variables are to be specified or measured. We are thus faced with (at least) two types of uncertainty, each of which brings about model uncertainty. Since there exist a wide set of possible model specifications, we often obtain contradictory conclusions. To make matters worse, estimation results are often not robust to small changes in model specification, making credible interpretations of the results hazardous. A proper treatment of model uncertainty is clearly important.

One such treatment is model averaging, where the aim of the investigator

is not to find the best possible model, but rather to find the best possible estimates. Each model contributes information about the parameters of interest, and all these pieces of information are combined taking into account the trust we have in each model, based on our prior beliefs and on the data.

In a sense, *all* estimation procedures are model averaging algorithms, although possibly extreme or limiting cases. Our framework is the linear regression model

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2), \quad (1)$$

where  $y$  ( $n \times 1$ ) is the vector of observations,  $X_1$  ( $n \times k_1$ ) and  $X_2$  ( $n \times k_2$ ) are matrices of nonrandom regressors,  $\varepsilon$  is a random vector of unobservable disturbances, and  $\beta_1$  and  $\beta_2$  are unknown parameter vectors. We assume that  $k_1 \geq 1$ ,  $k_2 \geq 0$ ,  $k := k_1 + k_2 \leq n - 1$ , that  $X := (X_1 : X_2)$  has full column-rank, and that the disturbances  $(\varepsilon_1, \dots, \varepsilon_n)$  are i.i.d.  $N(0, \sigma^2)$ .

The reason for distinguishing between  $X_1$  and  $X_2$  is that  $X_1$  contains explanatory variables which we want in the model on theoretical or other grounds (irrespective of the found  $t$ -ratios of the  $\beta_1$ -parameters), while  $X_2$  contains additional explanatory variables of which we are less certain. The columns of  $X_1$  are called ‘focus’ regressors, and the columns of  $X_2$  ‘auxiliary’ regressors.<sup>1</sup>

There are  $k_2$  components of  $\beta_2$ , and a different model arises whenever a different subset of the  $\beta_2$ ’s is set equal to zero. If  $k_2 = 0$ , then no model selection takes place. If  $k_2 = 1$ , then there are two models to consider: the unrestricted and the restricted model. If  $k_2 = 2$ , there are four models: the unrestricted, two partially restricted (where one of the two  $\beta_2$ ’s is zero), and the restricted model. In general, there are  $2^{k_2}$  models to consider. We denote the  $i$ -th model by  $\mathcal{M}_i$ , which we write as

$$y = X_1\beta_1 + X_{2i}\beta_{2i} + \varepsilon,$$

where  $X_{2i}$  denotes an  $n \times k_{2i}$  matrix containing a subset of  $k_{2i}$  columns of  $X_2$ , and  $\beta_{2i}$  denotes the corresponding  $k_{2i} \times 1$  subvector of  $\beta_2$ . We have of course  $0 \leq k_{2i} \leq k_2$ .

Model averaging estimation proceeds in two steps. In the first step we ask how to estimate the parameters, conditional upon a selected model. In

---

<sup>1</sup>Maybe Leamer (1978, p. 194) was the first to categorize variables into two classes, which he called ‘focus’ and ‘doubtful’, so that the focus variables are always in the model, while the doubtful variables can be combined in an arbitrary linear manner, a special case of which is exclusion. Later Leamer (1985, p. 309) preferred the use of ‘free’ instead of ‘focus’, because it is not always the case that the focus variables are the focus of a study — they are just the variables that are always in the equation.

the second step we compute the estimator as a weighted average of these conditional estimators. There exist both Bayesian and non-Bayesian ideas about how to estimate and how to find the weights. Our emphasis will be on the Bayesian framework; for the non-Bayesian approach, see Claeskens and Hjort (2003), Hjort and Claeskens (2003), Hansen (2007), and Liang et al. (2008).

The *unrestricted* estimator simply sets the weight for the unrestricted model (no restrictions on  $\beta_1$  or  $\beta_2$ ) to one and performs a single estimation. Similarly, the *restricted* estimator sets  $\beta_2$  to zero and estimates the resulting restricted model. Both estimators are, admittedly trivial, examples of a model averaging procedure. More interesting, and more common, are *general-to-specific* (GtS) estimators which do involve a model selection procedure, typically based on the ‘significance’ of parameters through their  $t$ -ratios. There are many problems with this procedure (see e.g. Magnus, 1999), but the most important is that the model selection procedure is completely separated from the estimation procedure. What is reported are therefore conditional estimates, but the researcher acts as if they are unconditional estimates. This problem is known as *pretesting*.

In order to combine model selection and estimation, the Bayesian method offers a natural framework. The basic equations of BMA were first presented by Leamer (1978, Sections 4.4–4.6), who proposed Bayesian averaging of *Bayesian* estimates. In the context of growth econometrics, BMA was first applied by Fernández et al. (2001a) and Brock and Durlauf (2001). BMA is flexible with respect to the size and exact specification of a model and it does not require the a priori selection of any model. Inference is based on a weighted average over all models. The idea of Bayesian averaging of *classical* estimates was first proposed in Raftery (1995) and later by Sala-i-Martin et al. (2004). In growth econometrics, BMA has proved useful, and recent applications include León-González and Montolio (2004), Sala-i-Martin et al. (2004), and Masanjala and Papageorgiou (2007). Recently, interest is growing in different aspects of growth empirics, such as nonlinearities, parameter heterogeneity, and endogeneity. BMA is also applied in other areas of economics; see for example Tsangarides et al. (2004), Crespo-Cuaresma and Doppelhofer (2007), Eicher et al. (2007a, 2007c), Masanjala and Papageorgiou (2008), and Prüfer and Tondl (2008). In short, BMA has become an important technique.

There are, however, two major problems with BMA. First, the computational burden is very substantial. In fact, it is usually impossible to get exact BMA estimates, in which case some MCMC method must be applied, of which the Metropolis-Hastings algorithm is the most common. Second, Bayesian techniques work well when prior information is available, in which

case they guide us as to how this information should be combined with information from the data. But when no prior information is available and nevertheless informative priors need to be specified (as is the case with BMA), then we need to reflect on the meaning and impact of these priors.

In addition to these two problems, there are some further uncomfortable aspects to BMA. One is that BMA takes different priors for the same parameter depending on which submodel is considered. This was also noted by Hjort and Claeskens (2003), and it is a little difficult to interpret. Another uncomfortable aspect is that — since exact BMA is computationally so demanding — it is very difficult to consider extensions, for example to nonspherical disturbances.<sup>2</sup>

Our proposed WALs method deals with all these problems. The computational burden is trivial, and the proposed prior is attractive because it is ‘neutral’ (mimicking ignorance) and also near-optimal in the sense of minimizing some risk or regret criterion (Magnus, 2002). It is based on the equivalence theorem of Magnus and Durbin (1999) and Danilov and Magnus (2004), and was originally developed to better understand pretesting.

The concept and treatment of ignorance is essential in both BMA and WALs. Suppose for simplicity that  $k_2 = 1$  in (1), so that there is only one auxiliary regressor  $x_2$  and only one auxiliary parameter  $\beta_2$ , and we have  $y = X_1\beta_1 + \beta_2x_2 + \varepsilon$  with  $\varepsilon \sim N(0, \sigma^2)$ . It is well-known that if we delete the auxiliary variable  $x_2$  from our regression equation, then  $R^2$  will always decrease, but  $\bar{R}^2$  (the adjusted  $R^2$ ) will decrease if, and only if, the  $t$ -ratio of the auxiliary parameter is smaller than one in absolute value. It is also well-known (Magnus and Durbin, 1999, Theorem 1) that if we define the ‘theoretical’  $t$ -ratio

$$\eta := \frac{\beta_2}{\sigma/\sqrt{x_2'M_1x_2}}, \quad M_1 := I_n - X_1(X_1'X_1)^{-1}X_1'$$

then  $\text{MSE}(\hat{\beta}_{1r}) \leq \text{MSE}(\hat{\beta}_{1u})$  if, and only if,  $|\eta| \leq 1$ , where  $\hat{\beta}_{1r}$  and  $\hat{\beta}_{1u}$  denote the restricted (with  $\beta_2 = 0$ ) and unrestricted estimator of  $\beta_1$  respectively. Hence we shall say that we are ‘ignorant’ (or ‘neutral’) about the auxiliary parameter  $\beta_2$  when (a) we don’t know whether  $\beta_2$  is positive or negative, and (b) we don’t know whether including the corresponding auxiliary regressor  $x_2$  will increase or decrease the mean squared error of the estimated focus parameter  $\beta_1$ . More formally, we choose the prior distribution in WALs such that the prior median of  $\eta$  is zero and the prior median of  $\eta^2$  is one.

---

<sup>2</sup>See, however, Doppelhofer and Weeks (2008) who study the robustness of BMA with respect to outliers and heteroskedasticity in the context of cross-country growth regressions. Magnus et al. (2009) extend WALs estimation to nonspherical disturbances.



This treatment of ignorance is further elaborated on and defended in Magnus (2002), and it is close to the idea in Masanjala and Papageorgiou (2008), who state that a posterior inclusion probability of 0.50 corresponds approximately to an absolute  $t$ -ratio of one. The proposed priors for WALS are taken from the Laplace distribution and thus generate bounded risk, in contrast to the normal prior adopted by BMA which generates unbounded risk. Figure 1 in Section 3.4 illustrates this essential difference.

In this paper we confront BMA with WALS, and apply both techniques to shed further light on the determinants of economic growth. In our growth estimations we use a set-up which allows us to distinguish between standard Solow growth determinants and determinants that have been suggested in so-called ‘new growth’ theories. Based on these analyses, we can not only draw conclusions on the most appropriate model averaging technique but also provide insights on the impacts of frequently used growth determinants.

The paper is organized as follows. The two main model averaging techniques are described in Sections 2 (BMA) and 3 (WALS). In Section 2 we extend the standard BMA theory to allow for the case where model selection takes place over a subset of the regressors. In Section 3 we extend the theory of WALS (developed in the context of pretesting), so that it can be used as a general model averaging technique. Sections 4–6 present the growth estimation set-up and results, and Section 7 concludes. An appendix contains a description and justification of our data and selected variables.

## 2 Bayesian model averaging (BMA)

The usual set-up for Bayesian model averaging is the special case of (1) where  $k_1 = 1$  and  $X_1 = \iota$  (the vector of ones), so that the constant term is present in all models and model selection takes place over all regressors except the constant term. Our treatment is more general and allows model selection to take place over a subset ( $X_2$ ) of the regressors, while the focus regressors (the columns of  $X_1$ ) are forced to be present in every model.

A very large literature exists on BMA, some of which is mentioned in the introduction. Useful literature summaries can be found in Raftery et al. (1997) and Hoeting et al. (1999).

### 2.1 Prior, likelihood, and posterior in model $\mathcal{M}_i$

Assuming that  $\mathcal{M}_i$  is the true model, the likelihood is given by

$$p(y | \beta_1, \beta_{2i}, \sigma^2, \mathcal{M}_i) \propto (\sigma^2)^{-n/2} \exp - \frac{S_i}{2\sigma^2}, \quad (2)$$

where  $S_i := (y - X_1\beta_1 - X_{2i}\beta_{2i})'(y - X_1\beta_1 - X_{2i}\beta_{2i})$ . Following standard Bayesian theory of the normal linear model (O'Hagan, 1994, Chapter 9), we impose the conventional improper prior distribution  $p(\sigma^2 | \mathcal{M}_i) \propto \sigma^{-2}$  together with a partially proper prior on  $\beta_1, \beta_{2i} | \sigma^2, \mathcal{M}_i$ :

$$p(\beta_1 | \sigma^2, \mathcal{M}_i) \propto 1, \quad \beta_{2i} | \beta_1, \sigma^2, \mathcal{M}_i \sim N(0, \sigma^2 V_{0i}),$$

where  $V_{0i}$  is a positive definite  $k_{2i} \times k_{2i}$  matrix to be specified later. The joint prior distribution is then

$$p(\beta_1, \beta_{2i}, \sigma^2 | \mathcal{M}_i) \propto (\sigma^2)^{-(k_{2i}+2)/2} \exp -\frac{\beta_{2i}' V_{0i}^{-1} \beta_{2i}}{2\sigma^2}. \quad (3)$$

To deal with partially proper (informative), partially improper (noninformative) priors is no trivial matter; see Bauwens et al. (1999, pp. 117–118). Our approach will be to think of the improper prior distribution as a special case of the following *proper* prior distribution:

$$p(\beta_1, \beta_{2i}, \sigma^2 | \mathcal{M}_i) \propto (\sigma^2)^{-(d_0+k_1+k_{2i}+2)/2} \exp -\frac{h_0\beta_1'\beta_1 + \beta_{2i}' V_{0i}^{-1} \beta_{2i} + a_0}{2\sigma^2}, \quad (4)$$

where the special case (3) occurs when  $h_0 = 0$ ,  $a_0 = 0$ , and  $d_0 = -k_1$ .<sup>3</sup>

Combining the prior (4) with the likelihood (2) gives the posterior

$$p(\beta_1, \beta_{2i}, \sigma^2 | y, \mathcal{M}_i) \propto (\sigma^2)^{-(d+k_1+k_{2i}+2)/2} \exp -\frac{R_i + a_i}{2\sigma^2}, \quad (5)$$

where  $d = d_0 + n$ ,

$$R_i := \begin{pmatrix} \beta_1 - b_{1i} \\ \beta_{2i} - b_{2i} \end{pmatrix}' V_i^{-1} \begin{pmatrix} \beta_1 - b_{1i} \\ \beta_{2i} - b_{2i} \end{pmatrix},$$

$$V_i^{-1} := \begin{pmatrix} X_1' X_1 + h_0 I_{k_1} & X_1' X_{2i} \\ X_{2i}' X_1 & X_{2i}' X_{2i} + V_{0i}^{-1} \end{pmatrix},$$

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} := V_i(X_1 : X_{2i})' y,$$

and

$$a_i := a_0 + y'y - y'(X_1 : X_{2i})V_i(X_1 : X_{2i})'y.$$

Hence the posterior density of  $\beta_1$ ,  $\beta_{2i}$ , and  $\sigma^2$  — given the data  $y$  and model  $\mathcal{M}_i$  — is the familiar normal-inverse-gamma distribution with parameters  $a_i$ ,  $d$ ,  $(b_{1i}, b_{2i})$ , and  $V_i$ .

---

<sup>3</sup>Whenever priors are used the question of sensitivity of the posterior moments to the priors is important. We do not examine this issue here. Recent examples of such prior robustness checks for BMA include Ley and Steel (2009) and Eicher et al. (2007b).

A little algebra gives

$$V_i = \begin{pmatrix} (X_1'X_1 + h_0I_{k_1})^{-1} + Q_iV_{2i}Q_i' & -Q_iV_{2i} \\ -V_{2i}Q_i' & V_{2i} \end{pmatrix}, \quad (6)$$

where

$$V_{2i}^{-1} = V_{0i}^{-1} + X_{2i}'M_1^*X_{2i}, \quad M_1^* = I_n - X_1(X_1'X_1 + h_0I_{k_1})^{-1}X_1',$$

and

$$Q_i = (X_1'X_1 + h_0I_{k_1})^{-1}X_1'X_{2i}.$$

From (6) we find

$$(X_1 : X_{2i})V_i(X_1 : X_{2i})' = I - M_1^* + M_1^*X_{2i}V_{2i}X_{2i}'M_1^*,$$

so that we can rewrite  $a_i$  as

$$\begin{aligned} a_i &= a_0 + y'y - y'(X_1 : X_{2i})V_i(X_1 : X_{2i})'y \\ &= a_0 + y'(M_1^* - M_1^*X_{2i}V_{2i}X_{2i}'M_1^*)y. \end{aligned}$$

We now specialize to the improper prior given in (3) by setting  $h_0 = 0$ ,  $a_0 = 0$ , and  $d_0 = -k_1$ . The matrix  $M_1^*$  then specializes to the idempotent matrix  $M_1 := I_n - X_1(X_1'X_1)^{-1}X_1'$ , and we have  $a_i = (M_1y)'A_i(M_1y)$ , where  $A_i := M_1 - M_1X_{2i}V_{2i}X_{2i}'M_1$ . We notice that  $a_i$  is a function of  $M_1y$  only and does not depend on  $X_1'y$ . It follows that

$$E(\beta_1 | y, \mathcal{M}_i) = b_{1i} = (X_1'X_1)^{-1}X_1'(y - X_{2i}b_{2i}), \quad (7)$$

$$E(\beta_{2i} | y, \mathcal{M}_i) = b_{2i} = (V_{0i}^{-1} + X_{2i}'M_1X_{2i})^{-1}X_{2i}'M_1y, \quad (8)$$

and, when  $n > k_1 + 2$ ,

$$\text{var}(\beta_1 | y, \mathcal{M}_i) = \frac{a_i}{n - k_1 - 2}((X_1'X_1)^{-1} + Q_iV_{2i}Q_i'), \quad (9)$$

$$\text{var}(\beta_{2i} | y, \mathcal{M}_i) = \frac{a_i}{n - k_1 - 2}V_{2i}. \quad (10)$$

## 2.2 Marginal likelihood of model $\mathcal{M}_i$

In order to find the marginal likelihood we return to the proper prior (4). Since  $|V_i| = |X_1'X_1 + h_0I_{k_1}|^{-1} \cdot |V_{2i}|$ , we obtain the marginal density of  $y$  in model  $\mathcal{M}_i$  as

$$\begin{aligned} p(y | \mathcal{M}_i) &= \iiint p(y | \beta_1, \beta_{2i}, \sigma^2, \mathcal{M}_i) p(\beta_1, \beta_{2i}, \sigma^2 | \mathcal{M}_i) d\beta_1 d\beta_{2i} d\sigma^2 \\ &= \frac{\pi^{-n/2} h_0^{k_1/2} a_0^{d_0/2} \Gamma(d/2)}{|X_1'X_1 + h_0I_{k_1}|^{1/2} \Gamma(d_0/2)} \cdot \frac{|V_{0i}^{-1}|^{1/2}}{|V_{2i}^{-1}|^{1/2}} \cdot a_i^{-d/2} \\ &= c \cdot |V_{2i}^{-1}|^{-1/2} |V_{0i}^{-1}|^{1/2} a_i^{-d/2}, \end{aligned}$$

where  $c$  is a normalizing constant which does not depend on  $i$  or  $y$ .

Now specializing to the improper prior (3) by setting  $h_0 = 0$ ,  $a_0 = 0$ , and  $d_0 = -k_1$ , we find

$$p(y | \mathcal{M}_i) = c \cdot \frac{|V_{0i}^{-1}|^{1/2}}{|V_{0i}^{-1} + X_{2i}' M_1 X_{2i}|^{1/2}} \cdot (y' M_1 A_i M_1 y)^{-(n-k_1)/2}, \quad (11)$$

where

$$A_i := M_1 - M_1 X_{2i} (V_{0i}^{-1} + X_{2i}' M_1 X_{2i})^{-1} X_{2i}' M_1$$

and  $M_1 = I_n - X_1 (X_1' X_1)^{-1} X_1'$ . If we let  $p(\mathcal{M}_i)$  denote the prior probability that  $\mathcal{M}_i$  is the true model, and  $\lambda_i := p(\mathcal{M}_i | y)$  the posterior probability for model  $\mathcal{M}_i$ , then

$$\lambda_i = \frac{p(\mathcal{M}_i) p(y | \mathcal{M}_i)}{\sum_j p(\mathcal{M}_j) p(y | \mathcal{M}_j)} \quad (i = 1, \dots, 2^{k_2}),$$

We shall assign equal prior probability to each model under consideration. This seems to be in line with the standard literature on BMA, although it is not without criticism and alternative choices for  $p(\mathcal{M}_i)$  have been proposed. Many researchers feel that simpler models should be preferred to more complex ones, all else being equal. Durlauf et al. (2005), on the other hand, find the idea of promoting parsimonious models through the priors unappealing. Brock and Durlauf (2001) raise objections against uniform priors on the model space because of the implicit assumption that the probability that one regressor appears in the model is independent of the inclusion of others, whereas, in fact, regressors are typically correlated. They suggest a hierarchical structure for the model prior. This, however, requires agreement on which regressors are proxies for the same theories. As stated in Eicher et al. (2007b), such an agreement is usually not within reach and, therefore, independent model priors seem a reasonable compromise. Thus motivated we write

$$p(\mathcal{M}_i) = 2^{-k_2}.$$

Then  $\lambda_i = p(y | \mathcal{M}_i)$ , where the normalizing constant  $c$  is chosen such that  $\sum_i \lambda_i = 1$ .

### 2.3 Model averaging

So far we have conditioned on one model, namely model  $\mathcal{M}_i$ . In the Bayesian framework it is now easy to consider all models in our assumed model space

$\mathcal{M} := \{\mathcal{M}_i, i = 1, \dots, 2^{k_2}\}$ , by writing the posterior distribution of our parameters  $\beta_1$ ,  $\beta_2$ , and  $\sigma^2$  given the data  $y$  as

$$p(\beta_1, \beta_2, \sigma^2 | y) = \sum_{i=1}^{2^{k_2}} \lambda_i p(\beta_1, \beta_{2i}, \sigma^2 | y, \mathcal{M}_i). \quad (12)$$

This is a weighted average of the posterior distributions under each model, weighted by the corresponding posterior model probabilities.

The posterior mean and variance of  $\beta_1$  are

$$b_1 := E(\beta_1 | y) = \sum_i \lambda_i b_{1i}, \quad (13)$$

and

$$\text{var}(\beta_1 | y) = \sum_i \lambda_i (V_{1i}^* + b_{1i} b_{1i}') - b_1 b_1', \quad (14)$$

where  $b_{1i} := E(\beta_1 | y, \mathcal{M}_i)$  and  $V_{1i}^* := \text{var}(\beta_1 | y, \mathcal{M}_i)$ ; see Raftery (1993) and Draper (1995).

To obtain the corresponding results for  $\beta_2$  we introduce the  $k_2 \times k_{2i}$  selection matrices  $T_i$  with full column-rank, so that  $T_i' = (I_{k_{2i}} : 0)$  or a column-permutation thereof, and  $T_i \beta_{2i}$  is the  $k_2 \times 1$  vector obtained from  $\beta_2$  by setting the components not included in  $\mathcal{M}_i$  to zero. The posterior mean and variance of  $\beta_2$  are then

$$b_2 := E(\beta_2 | y) = \sum_i \lambda_i T_i b_{2i}, \quad (15)$$

and

$$\text{var}(\beta_2 | y) = \sum_i \lambda_i T_i (V_{2i}^* + b_{2i} b_{2i}') T_i' - b_2 b_2', \quad (16)$$

where  $b_{2i} := E(\beta_{2i} | y, \mathcal{M}_i)$  and  $V_{2i}^* := \text{var}(\beta_{2i} | y, \mathcal{M}_i)$ .

## 2.4 Implementation using $g$ -priors

Following Zellner (1986) we assume that the prior variance  $V_{0i}$  is given by

$$V_{0i}^{-1} = g_i X_{2i}' M_1 X_{2i} \quad (g_i > 0).$$

This gives

$$\lambda_i = c \cdot \left( \frac{g_i}{1 + g_i} \right)^{k_{2i}/2} (y' M_1 A_i M_1 y)^{-(n-k_1)/2},$$

where

$$A_i = \frac{g_i}{1 + g_i} M_1 + \frac{1}{1 + g_i} (M_1 - M_1 X_{2i} (X_{2i}' M_1 X_{2i})^{-1} X_{2i}' M_1).$$

We have

$$b_{1i} = (X_1'X_1)^{-1}X_1'(y - X_{2i}b_{2i}),$$

$$b_{2i} = \frac{1}{1 + g_i}(X_{2i}'M_1X_{2i})^{-1}X_{2i}'M_1y.$$

Also, when  $n > k_1 + 2$  and defining  $s_i^2 := y'M_1A_iM_1y/(n - k_1 - 2)$ , we find

$$V_{1i}^* = s_i^2(X_1'X_1)^{-1} + (X_1'X_1)^{-1}X_1'X_{2i}V_{2i}^*X_{2i}'X_1(X_1'X_1)^{-1},$$

$$V_{2i}^* = \frac{s_i^2}{1 + g_i}(X_{2i}'M_1X_{2i})^{-1}.$$

Our final ingredient is the specification of  $g_i$ . We follow Fernández et al. (2001b) and choose

$$g_i := \frac{1}{\max(n, k_2^2)},$$

where we note that  $g_i$  is the same for all  $i$ . One alternative would have been  $g_i := 1/n$ , the so-called ‘unit information prior’ (Raftery, 1995), recently advocated by Eicher et al. (2007b), and Masanjala and Papageorgiou (2008). In our case with  $n = 74$  and  $k_2 = 4, 9$ , or  $12$ , the difference between the two priors is negligible.

The above results now allow us to calculate the BMA estimates and precisions of  $\beta_1$  and  $\beta_2$  from (13)–(16). Special cases arise and some care is required when  $k_2 = 0$  (no model selection) or  $k_1 = 0$  (model selection takes place over all regressors). Our Matlab program, downloadable from

<http://center.uvt.nl/staff/magnus/wals>,

allows for these special cases.

### 3 Weighted-average least squares (WALS)

#### 3.1 Orthogonalization

Weighted-average least squares estimation starts with the realization that we can ‘orthogonalize’ the columns of  $X_2$  such that  $X_2'M_1X_2 = I_{k_2}$ , where we recall that  $M_1 := I_n - X_1(X_1'X_1)^{-1}X_1'$ . More precisely, if we let  $P$  be an orthogonal  $k_2 \times k_2$  matrix such that  $P'X_2'M_1X_2P = \Lambda$  (diagonal), and define new auxiliary regressors  $X_2^* := X_2P\Lambda^{-1/2}$  and new auxiliary parameters  $\beta_2^* = \Lambda^{1/2}P'\beta_2$ , then  $X_2^*\beta_2^* = X_2\beta_2$  and  $X_2^{*'}M_1X_2^* = I_{k_2}$ . There are major advantages in working with  $X_2^*$  and  $\beta_2^*$  instead of  $X_2$  and  $\beta_2$ , as will become clear shortly. Hence, we shall *initially* assume that this orthogonalization

has taken place.

**Assumption 1:**  $X_2' M_1 X_2 = I_{k_2}$ .

Assumption 1 thus requires that the columns  $x_{21}, \dots, x_{2k_2}$  of  $X_2$  (the auxiliary regressors) are ‘orthogonal’ in the sense that  $M_1 x_{2i}$  and  $M_1 x_{2j}$  are orthogonal for every  $i \neq j$ . This will not affect the interpretation of the  $\beta_1$ -coefficients, but it will change the interpretation of the  $\beta_2$ -coefficients. However, we can always recover  $\beta_2$  from  $\beta_2 = P\Lambda^{-1/2}\beta_2^*$ .

### 3.2 Restricted least squares

Given Assumption 1, the least-squares (LS) estimators of  $\beta_1$  and  $\beta_2$  in the unrestricted model (1) are

$$\hat{\beta}_1 = \hat{\beta}_{1r} - Q\hat{\beta}_2, \quad \hat{\beta}_2 = X_2' M_1 y,$$

where  $\hat{\beta}_{1r} := (X_1' X_1)^{-1} X_1' y$  and  $Q := (X_1' X_1)^{-1} X_1' X_2$ . The subscript ‘ $r$ ’ denotes ‘restricted’ (with  $\beta_2 = 0$ ). We see that  $\hat{\beta}_2 \sim N(\beta_2, \sigma^2 I_{k_2})$ .

Let  $S_i$  be an  $k_2 \times (k_2 - k_{2i})$  selection matrix with full column-rank, where  $0 \leq k_{2i} \leq k_2$ , so that  $S_i' = (I_{k_2 - k_{2i}} : 0)$  or a column-permutation thereof. We are interested in the restricted LS estimators of  $\beta_1$  and  $\beta_2$ , the restriction being  $S_i' \beta_2 = 0$ . Let  $\mathcal{M}_i$  denote the linear model (1) under the restriction  $S_i' \beta_2 = 0$ , and denote the LS estimators of  $\beta_1$  and  $\beta_2$  in model  $\mathcal{M}_i$  by  $\hat{\beta}_{1i}$  and  $\hat{\beta}_{2i}$ . Following Danilov and Magnus (2004, Lemmas A1 and A2), the restricted LS estimators of  $\beta_1$  and  $\beta_2$  are given by

$$\hat{\beta}_{1i} = \hat{\beta}_{1r} - QW_i \hat{\beta}_2, \quad \hat{\beta}_{2i} = W_i \hat{\beta}_2, \quad (17)$$

where  $W_i := I_{k_2} - S_i S_i'$  is a *diagonal*  $k_2 \times k_2$  matrix with  $k_{2i}$  ones and  $(k_2 - k_{2i})$  zeros on the diagonal, such that the  $j$ -th diagonal element of  $W_i$  is zero if  $\beta_{2j}$  is restricted to be zero, and one otherwise. (If  $k_{2i} = k_2$  then  $W_i := I_{k_2}$ .) The joint distribution of  $\hat{\beta}_{1i}$  and  $\hat{\beta}_{2i}$  is then

$$\begin{pmatrix} \hat{\beta}_{1i} \\ \hat{\beta}_{2i} \end{pmatrix} \sim N_k \left( \begin{pmatrix} \beta_1 + QS_i S_i' \beta_2 \\ W_i \beta_2 \end{pmatrix}, \sigma^2 \begin{pmatrix} (X_1' X_1)^{-1} + QW_i Q' & -QW_i \\ -W_i Q' & W_i \end{pmatrix} \right),$$

the residual vector  $e_i := y - X_1 \hat{\beta}_{1i} - X_2 \hat{\beta}_{2i}$  is given by  $e_i = D_i y$ , where  $D_i := M_1 - M_1 X_2 W_i X_2' M_1$  is a symmetric idempotent matrix of rank  $n - k_1 - k_{2i}$ , and the distribution of  $s_i^2 := e_i' e_i / (n - k_1 - k_{2i})$  is

$$\frac{(n - k_1 - k_{2i}) s_i^2}{\sigma^2} \sim \chi^2(n - k_1 - k_{2i}, \frac{\beta_2' S_i S_i' \beta_2}{\sigma^2}).$$

It follows that:

- all models which include  $x_{2j}$  as a regressor will have the *same* estimator of  $\beta_{2j}$ , namely  $\hat{\beta}_{2j}$ , irrespective which other  $\beta_2$ 's are estimated;
- the estimators  $\hat{\beta}_{21}, \hat{\beta}_{22}, \dots, \hat{\beta}_{2k_2}$  are independent;
- if  $\sigma^2$  is known or is estimated by  $s^2$  (the LS estimator in the unrestricted model), then all models which include  $x_{2j}$  as a regressor yield the same  $t$ -ratio of  $\beta_{2j}$ .

### 3.3 The equivalence theorem

We now define the WALS estimator of  $\beta_1$  as

$$b_1 = \sum_{i=1}^{2^{k_2}} \lambda_i \hat{\beta}_{1i}, \quad (18)$$

where the sum is taken over all  $2^{k_2}$  different models obtained by setting a subset of the  $\beta_2$ 's equal to zero, and the  $\lambda_i$  are weight-functions satisfying certain minimal regularity conditions, namely

$$\lambda_i \geq 0, \quad \sum_i \lambda_i = 1, \quad \lambda_i = \lambda_i(M_1 y). \quad (19)$$

The WALS estimator can then be written as  $b_1 = \hat{\beta}_{1r} - QW\hat{\beta}_2$ , where  $W := \sum_i \lambda_i W_i$ . Notice that, while the  $W_i$  are nonrandom,  $W$  is random. For example, when  $k_2 = 2$ , we have four models to compare: the restricted  $\mathcal{M}_0$  ( $\beta_{21} = \beta_{22} = 0$ ), the partially restricted  $\mathcal{M}_1$  ( $\beta_{22} = 0$ ) and  $\mathcal{M}_2$  ( $\beta_{21} = 0$ ), and the unrestricted  $\mathcal{M}_{12}$ . The corresponding  $W_i$  are

$$W_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad W_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad W_{12} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and hence

$$W = \begin{pmatrix} \lambda_1 + \lambda_{12} & 0 \\ 0 & \lambda_2 + \lambda_{12} \end{pmatrix},$$

where the weight-functions  $\lambda_i$  are now labeled  $\lambda_0, \lambda_1, \lambda_2, \lambda_{12}$ , corresponding to the four models  $\mathcal{M}_i$  and matrices  $W_i$ . We see that  $\lambda_0$  does not appear in the matrix  $W$  and that  $W$  is diagonal because of Assumption 1.

A few words about the regularity conditions are in order. If  $\sigma^2$  is known, then most or all diagnostics will use statistics (such as  $t$ - and  $F$ -statistics) which depend on  $\hat{\beta}_2$  only. If  $\sigma^2$  is not known and estimated by  $s^2$ , then all  $t$ - and  $F$ -statistics will depend on  $(\hat{\beta}_2, s^2)$ . Now, it is a basic result in



least-squares theory that  $s^2$  is independent of  $(\hat{\beta}_1, \hat{\beta}_2)$ . It follows that  $\hat{\beta}_{1r}$  is independent of  $s^2$ . Hence,  $\hat{\beta}_{1r}$  will be independent of  $(\hat{\beta}_2, s^2)$ . Finally, if  $\sigma^2$  is not known and estimated by  $s_i^2$  (the estimator of  $\sigma^2$  in model  $\mathcal{M}_i$ ), then it is no longer true that all  $t$ - and  $F$ -statistics depend only on  $(\hat{\beta}_2, s^2)$ . However, they still depend only on  $M_1 y$ , because we have seen that both  $\hat{\beta}_{2i}$  and the residuals  $e_i$  from model  $\mathcal{M}_i$  are linear functions of  $M_1 y$ . We conclude that the regularity conditions on  $\lambda_i$  are reasonable and mild.

The equivalence theorem proved in Danilov and Magnus (2004, Theorem 1), generalizing an earlier result in Magnus and Durbin (1999), states that if Assumption 1 holds and the regularity conditions (19) on  $\lambda_i$  are satisfied, then

$$E(b_1) = \beta_1 - Q E(W\hat{\beta}_2 - \beta_2), \quad \text{var}(b_1) = \sigma^2(X_1'X_1)^{-1} + Q \text{var}(W\hat{\beta}_2)Q',$$

and hence

$$\text{MSE}(b_1) = \sigma^2(X_1'X_1)^{-1} + Q\text{MSE}(W\hat{\beta}_2)Q'. \quad (20)$$

The importance of the equivalence theorem lies in the fact that the properties of the complicated WALS estimator  $b_1$  of  $\beta_1$  depend critically on the properties of the less complicated estimator  $W\hat{\beta}_2$  of  $\beta_2$ . We notice that neither the bias, nor the variance or the mean squared error of  $b_1$  depend on  $\beta_1$ . They do, however, depend on  $\beta_2$ .

It follows from the equivalence theorem (from (20) in particular) that the WALS estimator  $b_1$  will be a ‘good’ estimator of  $\beta_1$  (in the mean squared error sense) if and only if  $W\hat{\beta}_2$  is a ‘good’ estimator of  $\beta_2$ . Now, under Assumption 1, the matrix  $W$  is diagonal, say  $W = \text{diag}(w_1, \dots, w_{k_2})$ . Suppose that  $\sigma^2$  is known (we discuss the unknown  $\sigma^2$  case later), and that we choose  $w_j = w_j(\hat{\beta}_{2j})$ . Then, since the  $\{\hat{\beta}_{2j}\}$  are independent, so are the  $\{w_j\hat{\beta}_{2j}\}$ , and our  $k_2$ -dimensional problem reduces to  $k_2$  (identical) one-dimensional problems: only using the information that  $\hat{\beta}_{2j} \sim N(\beta_{2j}, \sigma^2)$  and assuming that  $\sigma^2$  is known, find the best (in the mean squared error sense) estimator of  $\beta_{2j}$ . The Laplace estimator discussed below solves this problem.

Suppose  $\tilde{\beta}_{2j}$  is the desired optimal estimator of  $\beta_{2j}$ . Then, letting  $\tilde{\beta}_2 := (\tilde{\beta}_{21}, \dots, \tilde{\beta}_{2k_2})'$ , the equivalence theorem directly gives us the optimal WALS estimator

$$b_1 = \hat{\beta}_{1r} - Q\tilde{\beta}_2,$$

with

$$E(b_1) = \beta_1 - Q E(\tilde{\beta}_2 - \beta_2), \quad \text{var}(b_1) = \sigma^2(X_1'X_1)^{-1} + Q \text{var}(\tilde{\beta}_2)Q'.$$

From a computational point of view, it is important to note that the number of required calculations is of order  $k_2$ , even though there are  $2^{k_2}$  models to

consider. This is so because we do not need all  $2^{k_2}$  individual  $\lambda$ 's; only  $k_2$  linear combinations are required, namely the diagonal elements of  $W$ ; see Leamer (1978, p. 154) for a related result in terms of principal components.

### 3.4 The Laplace estimator

Thus motivated, let  $x$  be a single observation from a univariate normal distribution with mean  $\eta$  and variance one, that is,  $x \sim N(\eta, 1)$ . How to estimate  $\eta$ ? This seemingly trivial question was addressed in Magnus (2002). We consider five candidates (there are more):

- the ‘usual’ estimator:  $t(x) = x$
- the ‘silly’ estimator:  $t(x) = 0$
- the pretest estimator:

$$t(x) = \begin{cases} 0 & \text{if } |x| \leq c \\ x & \text{if } |x| > c \end{cases}$$

- the ‘normal’ estimator:  $t(x) = x/(1 + c)$
- the Laplace estimator defined below in (21),

where  $c$  is a (generic) nonnegative constant. The five estimators are graphed in Figure 1, where  $c = 1.96$  for the pretest estimator,  $c = 1/2.1981$  for the ‘normal’ estimator, and  $c = \log(2)$  for the Laplace estimator.

FIGURE 1

The usual estimator is unbiased, admissible, and minimax. Its risk  $R(\eta) := E(t(x) - \eta)^2 = 1$  has good properties when  $|\eta|$  is large, but not when  $\eta$  is close to zero. The silly estimator has excellent properties when  $\eta$  is close to zero, but its risk  $R(\eta) = \eta^2$  increases without bound when  $|\eta|$  becomes large. The pretest estimator has bounded risk, but it has a discontinuity and is therefore inadmissible. Also, its risk is higher than either the usual or the silly estimator when  $|\eta|$  is around one. The ‘normal’ estimator is a Bayesian estimator, combining the likelihood  $x|\eta \sim N(\eta, 1)$  with a normal prior  $\pi(\eta) \sim N(0, 1/c)$ . (In Figure 1 we take  $1/c = 2.1981$ , so that  $\Pr(|\eta| < 1) = 1/2$ .) This is — in essence — the BMA estimator. The risk of the ‘normal’ estimator is also unbounded. The Laplace estimator was developed as an estimator which is admissible, has bounded risk, has good properties around  $|\eta| = 1$ , and is

near-optimal in terms of minimax regret. It is a Bayesian estimator, based on the Laplace prior

$$\pi(\eta) = \frac{c}{2} \exp(-c|\eta|).$$

The hyperparameter  $c$  is chosen  $c = \log 2$ , because this implies that the prior median of  $\eta$  is zero and the prior median of  $\eta^2$  is one, which comes closest to our prior idea of ignorance as discussed in the Introduction.

The moments of the posterior distribution of  $\eta|x$  are given in Theorem 1, which extends Pericchi and Smith (1992) and Magnus (2002).

**Theorem 1:** Consider the likelihood and prior

$$x|\eta \sim N(\eta, 1), \quad \pi(\eta) = \frac{c}{2} \exp(-c|\eta|),$$

where  $c$  is a positive hyperparameter. Let  $Q(x, \eta) := (x - \eta)^2 + 2c|\eta|$ . Then the posterior distribution of  $\eta$  given  $x$  is given by

$$p(\eta | x) = \frac{\exp(-Q(x, \eta)/2)}{\int \exp(-Q(x, \eta)/2) d\eta}.$$

The mean and variance of the posterior distribution are given by

$$E(\eta | x) = \frac{1 + h(x)}{2}(x - c) + \frac{1 - h(x)}{2}(x + c) \quad (21)$$

and

$$\text{var}(\eta | x) = 1 + c^2(1 - h^2(x)) - \frac{c(1 + h(x))\phi(x - c)}{\Phi(x - c)},$$

where

$$h(x) := \frac{e^{-cx}\Phi(x - c) - e^{cx}\Phi(-x - c)}{e^{-cx}\Phi(x - c) + e^{cx}\Phi(-x - c)},$$

and  $\phi(x)$  and  $\Phi(x)$  denote the density and cumulative distribution function of the standard-normal distribution, respectively.

**Proof:** Writing

$$Q(x, \eta) = \begin{cases} (\eta - (x + c))^2 - 2cx - c^2 & \text{if } \eta \leq 0, \\ (\eta - (x - c))^2 + 2cx - c^2 & \text{if } \eta > 0, \end{cases}$$

and realizing that

$$\int_{-\infty}^x t\phi(t) dt = -\phi(x), \quad \int_{-\infty}^x t^2\phi(t) dt = \Phi(x) - x\phi(x),$$

the proof follows easily.  $\parallel$

The moments are easy to compute. Note that the function  $h$  is monotonically increasing with  $h(-\infty) = -1$ ,  $h(0) = 0$ , and  $h(\infty) = 1$ , and that  $h(-x) = -h(x)$ .

### 3.5 Implementation using Laplace priors

The WALS estimation procedure can be summarized as follows.

- In the unrestricted model  $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ , determine which are the focus regressors  $X_1$  and which are the auxiliary regressors  $X_2$ .
- Compute  $M_1 := I_n - X_1(X_1'X_1)^{-1}X_1'$ , and then  $P$  (orthogonal) and  $\Lambda$  (diagonal) such that  $P'X_2'M_1X_2P = \Lambda$ . Compute  $X_2^* := X_2P\Lambda^{-1/2}$ , so that  $X_2^{*'}M_1X_2^* = I_{k_2}$ . Letting  $\beta_2^* := \Lambda^{1/2}P'\beta_2$ , note that  $X_2^{*'}\beta_2^* = X_2\beta_2$ .
- Compute  $\hat{\beta}_2^* = X_2^{*'}M_1y$ .
- Let  $\eta := \beta_2^*/\sigma$ . Assuming that  $\sigma^2$  is known, compute  $\hat{\eta} := \hat{\beta}_2^*/\sigma$ . Notice that the components  $\hat{\eta}_1, \dots, \hat{\eta}_{k_2}$  of  $\hat{\eta}$  are independent and that  $\hat{\eta}_j \sim N(\eta_j, 1)$ .
- For  $j = 1, \dots, k_2$  compute the Laplace estimator  $\tilde{\eta}_j := E(\eta_j|\hat{\eta}_j)$  and its variance  $\omega_j^2 := \text{var}(\eta_j|\hat{\eta}_j)$ . Define  $\tilde{\eta} := (\tilde{\eta}_1, \dots, \tilde{\eta}_{k_2})'$  and  $\Omega := \text{diag}(\omega_1^2, \dots, \omega_{k_2}^2)$ .
- Since  $\eta = \beta_2^*/\sigma = \Lambda^{1/2}P'\beta_2/\sigma$ , we obtain  $\beta_2 = \sigma P\Lambda^{-1/2}\eta$ , and hence we compute the WALS estimators for  $\beta_2$  and  $\beta_1$  as

$$b_2 = \sigma P\Lambda^{-1/2}\tilde{\eta}, \quad b_1 = (X_1'X_1)^{-1}X_1'(y - X_2b_2).$$

- Letting  $Q := (X_1'X_1)^{-1}X_1'X_2$ , the variance of  $b_2$  and  $b_1$  is

$$\text{var}(b_2) = \sigma^2 P\Lambda^{-1/2}\Omega\Lambda^{-1/2}P',$$

and

$$\text{var}(b_1) = \sigma^2(X_1'X_1)^{-1} + Q\text{var}(b_2)Q'.$$

We also have  $\text{cov}(b_1, b_2) = -Q\text{var}(b_2)$ . In standard applications one is primarily interested in the diagonal elements of the variance matrices.

Finally, we note that we have assumed that  $\sigma^2$  is known, whereas in fact it is of course not known. Our solution to this problem is to replace  $\sigma^2$  by  $s^2$ , the estimate in the unrestricted model. This is an approximation, but a very accurate one, as demonstrated and exemplified by Danilov (2005).

### 3.6 BMA and WALS compared

It may seem at first glance that the two estimation procedures BMA and WALS are quite different, but in fact they are conceptually quite close. Both procedures are model averaging algorithms. The assumption that the data are normally distributed is the same, and the treatment of the focus parameters  $\beta_1$  and the error variance  $\sigma^2$  as noninformative priors is essentially the same. The difference between BMA and WALS lies in the prior treatment of the auxiliary parameters  $\beta_2$ . In BMA we assume normality of the priors with

$$E(\beta_{2i} | \mathcal{M}_i) = 0, \quad \text{var}(\beta_{2i} | \mathcal{M}_i) = \frac{\sigma^2}{g} (X'_{2i} M_1 X_{2i})^{-1},$$

where  $g := 1/\max(n, k_2^2)$ . Since  $\beta_{2i} = T'_i \beta_2$ ,  $X_{2i} = X_2 T_i$ , and  $T'_i T_i = I_{k_{2i}}$ , we can write these moments as

$$E(\beta_2 | \mathcal{M}_i) = 0, \quad \text{var}(\beta_2 | \mathcal{M}_i) = \frac{\sigma^2}{g} T_i (T'_i X'_2 M_1 X_2 T_i)^{-1} T'_i. \quad (22)$$

In contrast, in WALS we write  $\beta_2$  in terms of  $\eta$  as  $\beta_2 = \sigma P \Lambda^{-1/2} \eta$ . The  $k_2$  components of  $\eta$  are i.i.d. according to a Laplace distribution

$$\pi(\eta_i) = \frac{c}{2} \exp(-c|\eta_i|), \quad c = \log 2.$$

This implies that each  $\eta_i$  is symmetrically distributed around zero, that the median of  $\eta_i^2$  is one, and that the variance of  $\eta_i$  is  $\sigma_\eta^2 = 2/c^2$ . This choice of prior moments is based on our idea of ignorance as a situation where we don't know whether the theoretical  $t$ -ratio is larger or smaller than one in absolute value. The prior moments of  $\beta_2$  are then given by

$$E(\beta_2) = 0, \quad \text{var}(\beta_2) = \sigma^2 \sigma_\eta^2 P \Lambda^{-1} P' = \frac{\sigma^2}{c^2/2} (X'_2 M_1 X_2)^{-1}. \quad (23)$$

Comparing (22) and (23) shows that these prior moments are in fact closely related, and suggests in addition a new value for  $g$  in BMA applications, namely  $g = c^2/2 = 0.24$ .

The conceptual differences are thus the distribution (Laplace versus normal), where Laplace has the advantage of leading to finite risk; and the choice of  $g$  as a scaling parameter for the prior variance.

## 4 Growth models

In the neoclassical growth model (Solow, 1956), growth around a steady state is determined by rates of physical capital accumulation, population growth,

and exogenous technological progress. The initial income of an economy is relevant for its transition path as countries with a lower initial income are expected to grow faster than richer countries. The ‘new growth’ theories seek to explain also the previously exogenous components of economic growth, which is why they are often called ‘endogenous’ growth models. A frequently used empirical model for growth regressions is the human capital-augmented neoclassical model (Mankiw et al., 1992), which regresses the average growth rate of GDP per capita on investment, the log of initial GDP per capita, the population growth rate, and a human capital variable.

The ‘Solow’ determinants derived from a neoclassical growth model are sometimes called ‘proximate’ determinants because they are thought to be the most established drivers of economic growth. The term ‘proximate’ also reflects the ease with which these determinants can be influenced by policy measures, thus emphasizing their importance for empirical research and policy advice. Recent literature advocates the view that these proximate determinants in turn depend on slow-moving ‘fundamental’ growth determinants such as a country’s geography, the quality of its institutions, the degree of fractionalization in its society, and its culture or religion; see Durlauf et al. (2008a) and references therein. Hall and Jones (1999) use a similar framework in which they distinguish between proximate causes of economic success (capital accumulation and productivity) and a more fundamental determinant which they name ‘social structure’. Accordingly, one can distinguish between proximate and fundamental growth *theories* (Durlauf et al., 2008b).

We seek to capture these different types of growth determinants (theories) in our empirical analysis. Thus we define two different sets of regressors, labeled  $X_1$  and  $X_2$ , somewhat in the spirit of Brock and Durlauf (2001) and Masanjala and Papageorgiou (2008). The set  $X_1$  contains the regressors which appear in every regression on theoretical or other grounds (irrespective of their statistical significance), the so-called ‘focus’ regressors. Typically, but not necessarily,  $X_1$  contains the constant term as one of its regressors. The additional controls in the regression, the so-called ‘auxiliary’ regressors, are contained in  $X_2$ . Their primary role is to improve the estimation of the focus regressors, although their estimates may be of independent interest. The distinction between focus and auxiliary regressors is helpful when one wants to understand the relationship between neoclassical and other new growth determinants. While the Solow variables appear in many empirical studies, thus serving as a baseline for growth analysis, it is not so clear which variables should be included as auxiliary regressors. The proximate (Solow) determinants are the variables of major interest in our analysis ( $X_1$ ) based on their prominent position in growth theory and growth empirics. The fundamental growth determinants mentioned above are included as an

additional set of regressors ( $X_2$ ) serving as controls of the standard growth models.<sup>4</sup>

We analyze two different model specifications: Model 1 and Model 2. In the notion of Durlauf et al. (2008b), we interpret Model 1 as a direct test of the proximate neoclassical growth theory against the fundamental new growth theories of institutions, geography, fractionalization, and religion. Model 2 deviates from the proximate versus fundamental classification, and tests the robustness of the endogenous growth model using the distinction between focus and auxiliary regressors.

In both models the dependent variable is GROWTH. In our data set, the average growth rate is 1.99% with a standard error 1.86. The regressors and their role as either focus or auxiliary are given in Table 1.

TABLE 1

Model 1 contains six focus regressors (including the constant term) and four auxiliary regressors. It is motivated by the neoclassical growth model and thus contains all Solow determinants as focus regressors ( $X_1$ ). These are: The initial capital stock of an economy (GDP60), measured as the log of GDP per capita in 1960. This represents the so-called convergence term of the Solow growth model and attempts to analyze whether poorer countries (those having lower initial income) actually grow faster than richer ones. Next, the 1960–1985 equipment investment share of GDP (EQUIPINV), which serves as a proxy for the stock of physical capital in the economy and reflects the importance of capital accumulation for the growth of an economy. Then two variables which represent human capital. To capture different facets of human capital, we include a direct measure, the total gross enrollment rate in primary schooling in 1960 (SCHOOL60), and also a proxy for non-educational human capital, the life expectancy at age zero, measured in 1960 (LIFE60). Both human capital variables are widely used proxies for the initial human capital stock in an economy and are expected to have a positive effect on productivity growth with life expectancy being the more robust regressor (Sala-i-Martin, 1997). Whenever possible, we use such initial values for our variables, reducing also the potential endogeneity problem in our growth regressions. Finally, the population growth rate between 1960 and

---

<sup>4</sup>Our set-up is distantly related to the empirical study by Levine and Renelt (1992) who include (as we do) a set of variables that appear in every regression. They distinguish, however, between *three* sets of variables with the aim of finding the widest range of coefficient estimates on the variables of interest that standard hypothesis tests do not reject, thus assessing the robustness of partial correlations between the per capita growth rate and various economic indicators.

1990 (DPOP), a proxy for the exogenous growth rate of labor assumed to foster productivity growth in the neoclassical model.

To test this neoclassical model (theory) and its proximate growth determinants we include the suggested fundamental growth determinants as auxiliary regressors. There is not only theoretical but also empirical support for these regressors; see Sala-i-Martin (1997), Fernández et al. (2001a), and Sala-i-Martin et al. (2004). We specify the following set of four auxiliary variables in  $X_2$ . First, a rule of law index (LAW), a measure of the importance of institutions, supposed to have a positive effect on economic growth. Next, a country's fraction of tropical area (TROPICS), which controls for the effect of geography and is expected to have a negative effect on productivity growth. Third, an average index of ethnolinguistic fragmentation in a country (AVELF), which will help to analyze the influence of the degree of fractionalization in society and culture on economic productivity, typically found to be negative. And finally, the fraction of Confucian population in a country (CONFUC), used as a (somewhat dubious) proxy for culture or religion, typically identified as having a positive effect on growth. CONFUC can also be viewed as a proxy for the 'Asian (baby) tigers': Hong Kong, Malaysia, Singapore, South Korea, and Taiwan. However, as not all of the Asian countries with large growth rates are Confucian, CONFUC is more than just a regional dummy.

Model 2 contains nine focus regressors and four auxiliary regressors, and it represents an endogenous growth model trying to identify more specifically the factors driving growth and technological progress than is possible in Model 1. All regressors of our first model are included in Model 2 as focus regressors, except DPOP which is now an auxiliary regressor, because of its ambiguous role in economic growth. This ambiguity and lack of robustness was found, for example, by Sala-i-Martin (1997), Fernández et al. (2001a), and Sala-i-Martin et al. (2004). Our results reported in Section 5 confirm this ambiguity.

The three new auxiliary regressors are: the fraction of GDP produced in mining (MINING), a structural variable supposed to exert a negative effect on economic growth; an index for political rights (PRIGHTS), serving as a second institutional variable (the other is LAW), so that we capture not only the quality of the legal framework in a country but also a notion of public participation in the political process; and malaria prevalence in 1966 (MALARIA), another geographical variable (next to TROPICS), so that we account not only for the geographical location of a country, but also for its disease environment.

For each of Models 1 and 2, we estimate two versions: Set-up 1 as described above, and Set-up 2, where only the constant is a focus variable and



all other variables (nine in Model 1 and twelve in Model 2) are auxiliary. Set-up 2 is the typical model averaging framework and allows us to relate our results directly to previous studies.

## 5 Estimation results

We thus have two models and for each model we have two set-ups. We write the four models as 1(1), 1(2), 2(1), and 2(2), respectively. For each of these four models we consider five methods of estimation:

- Unrestricted: No model selection takes place. We estimate the model with all focus and all auxiliary regressors by OLS.
- Restricted: No model selection. We estimate the model with all focus regressors and none of the auxiliary regressors, also by OLS.
- GtS: General-to-specific model (GtS) selection takes place over the auxiliary regressors using Matlab's *stepwisefit* routine. In our version of GtS we start with the unrestricted model, then 'go down' (remove regressors), but never go 'back up' (add regressors). The values of the controls (`penter` =  $10^{-10}$  and `premove` = 0.05) reflect this choice. The selected model thus contains all focus regressors and a subset of the auxiliary regressors. This selected model is then estimated by OLS without pretesting taking into account. The reported OLS estimates and standard errors are thus conditional on the model selected.
- WALs: Weighted-average least squares estimation as discussed in Section 3.
- BMA: Bayesian model averaging as discussed in Section 2.

We note in passing that the estimates and standard errors reported for WALs and BMA are *not* conditional on inclusion. Some authors present the posterior moments conditional on inclusion (Sala-i-Martin et al., 2004) or they present both conditional and unconditional moments (Ley and Steel, 2007). While the conditional moments certainly contain information of interest, the unconditional moments are the ones that should be reported. To see why, let us consider the simplest case  $y_i = \alpha + \beta x_i + \varepsilon_i$ , where the constant term is a focus regressor and  $x$  is auxiliary, so that there are two models to consider: one where  $\beta = 0$  and one where  $\beta$  is to be estimated. Suppose that  $\bar{x} = 0$ . Then the unrestricted model gives  $\hat{\beta}_u = \sum x_i y_i / \sum x_i^2$ , while the restricted model gives  $\hat{\beta}_r = 0$ . The estimator for  $\beta$  is a weighted average of

these two:  $\hat{\beta} = \lambda\hat{\beta}_u + (1 - \lambda)\hat{\beta}_r = \lambda\hat{\beta}_u$ , where  $\lambda$  is determined by priors and data. The estimator  $\hat{\beta}_u$  is the estimator conditional on inclusion, and, while its moments contain information of interest, it will overestimate the impact of  $x$  on  $y$ , which is correctly estimated by the (unconditional) estimator  $\hat{\beta}$ . Hence, the unconditional moments are the ones that should be reported.

#### TABLES 2–5

The estimation results for the four models and five estimation methods are given in Tables 2–5. All regressors have the same signs across our estimation methods, model specifications, and set-ups with one exception, namely DPOP in the restricted estimation method of Model 1(1), which has a negative sign rather than the positive sign expected from neoclassical theory. We note that the standard errors are very large for this parameter, and that its sign (and its value) is therefore statistically not robust. In fact, its standard error is by far the largest of all regressors in all our estimations. The regressor PRIGHTS also enters with an unexpected negative sign in all estimations of Model 2, and does not seem to be a robust regressor either. This resembles the results of Sala-i-Martin et al. (2004), and could be due to the fact that most of the potentially beneficial effects of political stability on a country’s economic growth performance are already captured by other variables in the estimation, most notably LAW.

The regressors GDP60, TROPICS, and AVELF in Models 1 and 2, and the regressors MINING and MALARIA in Model 2 are negatively correlated with growth, which is reasonable, because a negative effect of initial GDP reflects (conditional) convergence between countries, and an unfavorable geographical location in the tropics (highly correlated with high rates of malaria prevalence) or a higher degree of fractionalization in a country are seen as impediments to economic growth. This is also the case for large endowments of natural resources which are usually associated with more political instability, rent-seeking, and low growth. All other regressors exhibit positive signs indicating that higher shares of physical and human capital, stability in terms of a sufficient rule of law, and a larger fraction of Confucian population foster economic growth. Not only the signs, but also the sizes of the estimates are closely correlated over the four models.

As the general-to-specific (GtS) model selection procedure is commonly used in practice, closer investigation of the selected models for 1(1), 1(2), 2(1), and 2(2) is appropriate. In Set-up 1 we have four auxiliary regressors, both in Models 1(1) and 2(1). In Model 1(1) the GtS procedure selects three of the four auxiliary regressors, while in Model 2(1) only one is selected. One would expect, perhaps, that in Model 1(1) the GtS estimates are close to the

unrestricted estimates, while in Model 2(1) they are close to the restricted estimates. This, however, is not the case. In Set-up 2 only the constant term is forced to be present in all models, while model selection takes place over all other variables. This is the procedure most commonly used. The GtS method selects only four of the nine regressors in Model 1(2), and six of the twelve regressors in Model 2(2). In Model 1(2), GtS drops two of the Solow determinants, namely SCHOOL60, and DPOP. This leads to a highly increased significance of the two other Solow determinants EQUIPINV and LIFE60. Also, the fundamental regressor CONFUC becomes much more important while the effect of GDP60 diminishes statistically and economically. Similar comments apply to Model 2(2). The statistical properties and conclusions of GtS must, however, be treated with caution, because — unlike WALS and BMA — the reported moments are conditional on the selected model, and the noise generated by the model selection procedure is ignored. This is the so-called pretesting problem, common to all classical model selection procedures.

Our main interest is in the comparison of the two averaging methods: BMA and WALS. The estimated coefficients seem to be somewhat higher (in absolute value) for WALS than for BMA (with the exception of EQUIPINV and CONFUC), especially for the auxiliary regressors, while the estimated standard errors are about the same on average. The economic impact of all robust and important regressors does not vary much between BMA and WALS. To shed some light on the relative importance of each regressor we compute the posterior inclusion probability (*pip*) and the *t*-ratio for each of the BMA estimates, and the *t*-ratio for each of the WALS estimates (since *pip* can not be computed for WALS) for Models 1(1) and 1(2).

TABLE 6

As a rough guideline for ‘robustness’ of a regressor, a value  $pip = 0.5$  is sometimes recommended (Raftery, 1995), corresponding approximately with an absolute *t*-ratio of  $|t| = 1$  (Masanjala and Papageorgiou, 2008); see our discussion on ignorance in the Introduction. We see from Table 6 that  $pip = 1$  for each of the focus regressors, because these regressors are in the model with probability one. The ordering of the four auxiliary regressors according to *pip* is the same as the ordering according to the *t*-ratios, both for BMA and for WALS. For the six focus regressors the *pip*-values are uninformative, and the correlation between the *t*-ratios of BMA and WALS is very high. Similar remarks apply to Model 2.

We see that CONFUC is by far the most robust auxiliary regressor with  $pip = 0.99$  and a WALS *t*-ratio of 2.72 in Model 1(1). LAW is the second

important auxiliary regressor with  $pip = 0.68$  and a WALS  $t$ -ratio of 2.25. In Model 1(2) the BMA estimates confirm the standard results that GDP60 ( $pip=0.98$ ) and EQUIPINV ( $pip = 0.88$ ) are the most important Solow determinants. Surprisingly, also DPOP seems important ( $pip = 0.85$ ). The order of importance is essentially the same when we consider WALS  $t$ -ratios, except that GDP60 ( $|t| = 4.16$ ) has the highest  $t$ -ratio followed by LIFE60 ( $|t| = 2.34$ ), while EQUIPINV ( $|t| = 1.93$ ) is only third. Among the fundamental determinants we find again that CONFUC and LAW are important variables, both in terms of  $pip$  and  $t$ -ratio.

The importance comparisons for Model 2 are similar. Among the robust regressors we find GDP60 ( $pip = 0.99$ ), CONFUC ( $pip = 0.97$ ), LAW ( $pip = 0.85$ ), EQUIPINV ( $pip = 0.83$ ), and LIFE60 ( $pip = 0.72$ ). Interestingly, MALARIA ( $pip = 0.87$ ) is the third most important regressor, reflecting the usually poor economic performance of tropical countries and, jointly with life expectancy, the large effects of health on economic outcomes. Again, the order of the most robust regressors is essentially the same whether we use BMA  $pip$ -values or WALS  $t$ -ratios.

Our model averaging results regarding the identification of the important regressors are mostly in line with the literature for both models and set-ups. We find some of the Solow or proximate determinants *not* among the robust and important regressors, notably the population growth rate, DPOP, and primary schooling, SCHOOL60, which is in line with other studies; see Fernández et al. (2001a), Sala-i-Martin et al. (2004), Eicher et al. (2007b), and Durlauf et al. (2008b). The fundamental determinant (theory) of fractionalization, AVELF, is not robust either, a finding confirmed by other model averaging studies (Fernández et al., 2001a; Sala-i-Martin et al., 2004; Durlauf et al., 2008b).

The recent study by Durlauf et al. (2008b) addresses the issue of theory robustness versus variable robustness in a two-stage least squares BMA framework with hierarchical priors. They do not find any of the fundamental theories geography, institutions, or religion to be robustly and directly correlated with growth, and they conjecture an *indirect* effect from institutions and religion on growth via proximate determinants. Religion is not robust as a theory, they claim, because, if dummies allowing for regional heterogeneity are included in the regression, the  $pip$ -value of religion drops sharply. Similarly, institutions are not robust as a theory, because only in estimations where solely the fundamental theories are present, institutions have a high  $pip$ -value.

While our results regarding religion and geography support the findings of Durlauf et al. (2008b), this is not the case regarding institutions. Our results provide evidence for a robust *direct* impact of institutions on economic

growth. The magnitude and robustness of our institutions variable, LAW, is for both models larger in Set-up 1 (which distinguishes between types of determinants or ‘theories’) than in Set-up 2 (which does not). This is especially interesting for Model 1 where LAW is ‘only’ included in the set of auxiliary regressors but nevertheless exerts a direct influence. Since the effect of institutions on growth is of obvious importance, we investigated this issue a little further. We added the same regional dummies and macroeconomic variables as in Durlauf et al. (2008b), and found that our institutions variable, LAW, is unaffected by these changes, irrespective whether we add the new variables to the set of focus regressors or to the set of auxiliary regressors. The *pip*-value is always larger than 0.9 and the absolute *t*-ratio always larger than 2.0. If anything, the inclusion of the additional variables strengthens the effect of LAW on growth.

## 6 WALS estimation of the full SDM data set

To gain further insight we also consider a much larger data set, namely the data analyzed by Sala-i-Martin et al. (2004), henceforth SDM. The SDM data set contains 88 countries and 68 regressors (67 explanatory variables plus the constant term). Estimates are presented in SDM (Table 2) using the so-called Bayesian averaging of classical estimates (BACE) approach. The SDM data have also been analyzed by Ley and Steel (2007).<sup>5</sup> The data used by Ley and Steel are the same as SDM except that the dependent variable has been multiplied by 100. Hence all estimates and standard errors must be divided by 100 to obtain comparable results.

To compare the results obtained by SDM and Ley and Steel (2007) with our approach, we estimate two versions. In WALS-*F1* there is only one focus regressor (the constant term) so that model selection takes place over 67 variables, as in SDM and Ley and Steel (2007). This corresponds to Set-up 2 of the application in Sections 4 and 5. In WALS-*F8* we select eight focus regressors (including the constant term), corresponding to Set-up 1 in Sections 4 and 5. We select the focus regressors in such a way that they resemble our focus regressors in Model 2.<sup>6</sup> The selected focus regressors are marked

---

<sup>5</sup>Ley and Steel (2007) present standardized results, but the nonstandardized results (which we need here) can be easily computed or downloaded from Mark Steel’s website: <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic/steel/steelhomepage/bma> by clicking on the link ‘Supplementary Material with Data and Fortran Code’ associated with Ley and Steel (2007), and extracting the file k67i9\_NStd.out. The results are in the table with heading ‘Betas: Posterior Moments (Unconditional and Conditional on Inclusion).’

<sup>6</sup>Instead of EQUIPINV, SDM employ IPRICE as the variable for domestic investment:

( $F$ ) in Tables 7 and 8. In both versions computing time is negligible.

TABLE 7

Table 7 gives the posterior means and standard errors resulting from three estimation methods using the full SDM data set: BACE, BMA, and WALS. As in SDM we present the posterior moments for BACE only conditional on inclusion, and, as in Ley and Steel (2007), we present both the unconditional and the conditional posterior moments for BMA, labeled BMA- $u$  and BMA- $c$ , respectively. The estimates and standard errors reported for WALS are *not* conditional on inclusion.

We first comment briefly on the signs and the magnitudes of the estimated coefficients over the different methods. Regarding the signs of the estimated coefficients we see from Table 7 that there are no sign changes between BMA- $u$  and BMA- $c$ , and that there is only one sign change (variable 67) between BACE and BMA. Hence these three columns are almost perfectly correlated in terms of signs. WALS- $F1$  and WALS- $F8$  are also highly correlated in terms of signs: there are only two sign changes (variables 08 and 35) between these two columns. However, BACE/BMA and WALS are less correlated: about 70–75% of the WALS estimates have the same sign as the corresponding BACE/BMA estimates.

It seems that the signs produced by WALS are more intuitive than those produced by BACE/BMA. For example, WALS finds positive correlation between being an European economy (variable 37) and growth, while BACE and BMA find negative correlation. Also, WALS finds that being under socialist rule (variable 52) has a negative impact on growth, while BACE and BMA report positive correlation.

We see that WALS- $F1$  differs twice from WALS- $F8$  in terms of sign, and in both cases the sign of WALS- $F8$  is more intuitive: a positive impact of life expectancy (variable 08) on growth and a negative impact of the fraction of Catholics (variable 35). Thus, the structure provided by distinguishing between focus and auxiliary regressors seems to help also in finding ‘correct’ signs.

Regarding the magnitudes of the estimated coefficients we see from Table 7 that the WALS estimates of the robust and important estimates are mostly in line with BACE and BMA, and with the literature in general. This is true for both versions of WALS, and it is also true for the identification of

---

the average investment price level between 1960 and 1994 on purchasing power parity basis. Source: Heston et al. (2001). A variable proxying for the rule of law is not included in the SDM data.

the important regressors.

Table 7 shows that the WALS estimates are somewhat higher (in absolute value) than the BMA and BACE estimates, especially the BMA-*u* estimates. This confirms and strengthens the results of Section 5. In addition, the estimated standard errors are much larger for WALS than for BACE and BMA, which demonstrates again the danger of using the conditional estimates: the standard errors are seriously underestimated.

TABLE 8

In order to emphasize the last point, we consider — in addition to the signs and magnitudes of the estimates — also their estimated precisions. Thus we present in Table 8 the absolute *t*-ratios for a subset of eleven regressors of particular interest.<sup>7</sup> We shall denote the *t*-ratio by  $t_c$  if it is calculated conditional on inclusion, and by  $t_u$  if it is not. Comparing the absolute *t*-ratios, we see that on average  $|t_c| = 2.36$  (BACE) and  $|t_c| = 2.59$  (BMA-*c*) for the conditional estimates, and  $|t_u| = 0.67$  (WALS-*F1*),  $|t_u| = 0.85$  (WALS-*F8*), and  $|t_u| = 0.82$  (BMA-*u*) for the unconditional estimates. In the case of Table 8 the  $t_c$ -ratios are more than three times as large (on average, absolute values) as the  $t_u$ -ratios. If we consider all 67 regressors (without the constant term), then the average *t*-ratios (in absolute value) are uniformly lower:  $|t_c| = 1.33$  (BACE) and  $|t_c| = 1.41$  (BMA-*c*) for the conditional estimates, and  $|t_u| = 0.41$  (WALS-*F1*),  $|t_u| = 0.44$  (WALS-*F8*) and  $|t_u| = 0.27$  (BMA-*u*) for the unconditional estimates, but the  $t_c$ -ratios are now almost four times as large (on average, absolute values) as the  $t_u$ -ratios.<sup>8</sup> This clearly demonstrates that conditional methods produce much more precise estimates than unconditional methods (such as WALS). However, this high precision is misleading and incorrect. It appears that the WALS estimates are in-between BMA-*c* and BMA-*u* in terms of precision.

We note from Table 8 that WALS-*F8* gives higher precisions than WALS-*F1* as one would expect, because it is based on more restrictions. This confirms the findings in Table 6 for the much smaller data set. It appears therefore that the more structured framework in WALS-*F8*, allowing for a distinction between focus and auxiliary regressors, also helps to better identify the most robust and important determinants (theories) of economic growth.

The previous analysis shows three things. First, the difference between

<sup>7</sup>The *t*-ratios are calculated from the original output using more than four decimal points, and not from Table 7.

<sup>8</sup>The average absolute *t*-ratio for BACE is calculated over 65 rather than 67 variables, because two of the estimates have a standard error of zero up to all decimal points available to us.

WALS and BMA- $u$  is not large. Second, the difference between conditional and unconditional estimates is very large, and the use of the conditional estimates gives unrealistically precise estimates. Given that SDM use only uniform priors in BACE, their high precisions (with a small data set and a large number of regressors) seem astonishing. As argued, the reason lies in the use of conditional instead of unconditional estimates. Third, the WALS results are mostly in line with the SDM and BMA results (certainly with the BMA- $u$  results, which are the more relevant) regarding the identification of the most robust and important determinants of economic growth. The WALS estimates have similar magnitudes and directions of influence, and where they differ it appears that the WALS estimates are often more intuitive. This shows that the WALS method can also be useful for very large data sets.

Given that WALS not only provides the posterior moments in negligible computing time, but also does this based on exact calculations without using an approximation algorithm (contrary to BACE and BMA), we believe that WALS is a suitable model averaging technique also (and in particular) for large data sets.

## 7 Conclusions

The presently available myriad of growth determinants exposes growth regressions to a high degree of model uncertainty. Solow's (1956) neoclassical growth model provides an important benchmark, but numerous other growth models have been proposed and estimated since 1956. Since estimates and policy recommendations based on a model without taking the model selection procedure explicitly into account can be seriously biased and are likely to underestimate the variance, it is important to develop estimation techniques that take model uncertainty explicitly into account in an integrated one-step procedure. BMA is one such procedure, and so is WALS. Rather than trying to find the best possible model (step one) and — conditional on the selected model — the best possible estimates (step two), it is usually more relevant to find the best possible estimates taking account of all information provided by all models (one step).

At present, BMA is a standard method in growth econometrics; it is flexible with respect to the size and exact specification of a model, and it does not require the a priori selection of any model. In this paper we confront BMA with WALS, a new method previously not used for (growth) estimations. WALS has a theoretical advantage over BMA in that it presents an explicit and transparent treatment of ignorance, and a practical advantage in that the required computing time is linear in the number of regressors rather than



exponential.

We apply these two model averaging techniques taking different types of growth determinants (theories) into account. We define two sets of regressors: focus regressors which we want in the model on theoretical or other grounds, and auxiliary regressors which contain additional explanatory variables of which we are less certain. The distinction between focus and auxiliary regressors is helpful when one wants to understand the relationship between neoclassical and other new growth determinants. While the Solow variables appear in many empirical studies, thus serving as a baseline for growth analysis, it is not so clear which variables should be included as auxiliary regressors.

Based on this framework we analyze two different model specifications, labeled Model 1 and Model 2. In the notion of Durlauf et al. (2008b), we interpret Model 1 as a direct test of the proximate neoclassical growth theory against the fundamental new growth theories of institutions, geography, fractionalization, and religion. Model 2 deviates from the proximate versus fundamental classification, and tests the robustness of the endogenous growth model using the distinction between focus and auxiliary regressors.

We also consider a much larger data set, namely the data analyzed by Sala-i-Martin et al. (2004). To compare their results and the estimates of Ley and Steel (2007) (using the same data) with our WALS approach we estimate two versions. In WALS-*F1* model selection takes place over all 67 variables, as in SDM and Ley and Steel (2007). In contrast, WALS-*F8* select eight focus regressors that resemble our focus regressors in Model 2. Both versions of WALS produce results in line with the other methods, while the computing time is negligible.

Our model averaging results regarding the magnitude of the robust and important estimates are mostly in line with the literature for both models and set-ups, and this is also true for the identification of the important regressors. Our results do, however, shed new light on the robustness of growth determinants and theories. In particular, we find that robust growth theories should include not only neoclassical growth variables but also institutions; and also that the choice of variables within the theory of institutions matters. Both findings are in contrast to the recent study by Durlauf et al. (2008b) which addresses the issue of theory robustness versus variable robustness.

The search for a robust growth theory continues to be a delicate venture. It seems advisable to address the robustness and the interplay of different types, such as proximate versus fundamental, at the level of growth *regressors* rather than at the level of growth *theories*, since model averaging procedures can produce quite different results with respect to the robustness of growth theories depending on the specific variables used to proxy them.

From an econometric theory point of view we conclude that WALS should be considered as a serious new model averaging technique, both theoretically and computationally. Simulation studies will need to provide further insights; see Magnus et al. (2009). A major advantage of WALS is that it is based on a transparent treatment of ignorance, while BMA will always depend on subjective (and possibly sensitive) choices of the hyperparameters, such as the specification of  $g_i$  in Section 2.4.

## Acknowledgements

We are grateful to Matteo Ciccarelli, Drew Creal, Dmitry Danilov, Gernot Doppelhofer, Steven Durlauf, Gary Koop, Ed Leamer, Eduardo Ley, Aart de Vos, Alan Wan, and Xinyu Zhang for helpful advice and discussions; to seminar participants at Tilburg University, CEMFI Madrid, the City University of Hong Kong, the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, the Università degli Studi di Roma Tor Vergata, the European Meeting of the Econometric Society in Milan (August 2008), and the North American Winter Meeting of the Econometric Society in San Francisco (January 2009) for constructive comments; and to the editor, associate editor and referee for their perceptive and helpful reports.

## Data Appendix

Our dependent variable (GROWTH) is the growth of per capita GDP between 1960 and 1996. Note that we wish to explain the growth rate in each country over a 37-year period, not the annual growth rates. Our data constitute a cross section of 74 countries worldwide, and for each country we require the observed growth rate over the 37-year period and observations of the relevant regressors. We confine ourselves to thirteen regressors, including the constant term.

The data used in Sections 4 and 5 are taken from the standard literature on growth regressions. Our primary source is the SDM data set (Sala-i-Martin et al., 2004) for 1960–1996, available on Gernot Doppelhofer’s website:

<http://www.econ.cam.ac.uk/faculty/doppelhofer/research/bace.htm>.

There are only two variables where we deviate from the SDM data, namely the equipment investment variable (EQUIPINV), which replaces another investment variable in SDM, and the rule of law index (LAW), which is not in the SDM data. Both variables are taken from Sala-i-Martin (1997) (henceforth SALA).

The SDM data set is based on a list of 139 countries. Since many of the regressors are not observed in each country, they selected 88 countries for their analysis. The SALA data set lists 134 countries (with many missing observations). The intersection of the two sets and our selected variables contains 74 countries, as follows:

**Africa** (18): Algeria, Botswana, Cameroon, Congo, Ethiopia, Ghana, Kenya, Madagascar, Malawi, Morocco, Nigeria, Senegal, Somalia, Tunisia, Uganda, Zaire, Zambia, Zimbabwe;

**Latin America & Caribbean** (21): Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Trinidad and Tobago, Uruguay, Venezuela;

**North America** (2): Canada, United States;

**Asia** (14): Bangladesh, Hong Kong, India, Israel, Japan, Jordan, Korea, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan, Thailand;

**Europe** (17): Austria, Belgium, Denmark, Finland, France, West Germany, Greece, Ireland, Italy, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom;

**Oceania** (2): Australia, Papua New Guinea.

The dependent variable is:

**GROWTH**: Growth of GDP per capita at purchasing power parity between 1960 and 1996 (base year 1996); calculated by Heston et al. (2001). Source: SDM.

Apart from the constant term, there are twelve regressors, briefly described as follows:

**GDP60**: Logarithm of GDP per capita in 1960; calculated by Heston et al. (2001). Source: SDM.

**EQUIPINV**: 1960–1985 real equipment investment share of GDP comprising producer's investments in electrical and nonelectrical machinery (measured in relative prices constant across countries); calculated by De Long and Summers (1991). Source: SALA.

**SCHOOL60**: Total gross enrollment ratio for primary education in 1960. Calculated by Barro and Lee (1993) from UNESCO data. Source: SDM.

**LIFE60**: Life Expectancy at age 0 in 1960. Calculated by Barro and Lee (1993) from World Development Reports, World Bank, and other national data sets. Source: SDM.

**DPOP:** Average growth rate of population between 1960 and 1990. Calculated by Barro and Lee (1993) from World Bank data. Source: SDM.

**LAW:** Index for the overall maintenance of the rule of law (also referred to as ‘law and order tradition’). Calculated by Knack and Keefer (1995) from the International Country Risk Guide. Source: SALA.

**TROPICS:** Proportion of a country’s land area within geographical tropics. Calculated by Gallup et al. (2001) from Arc World Supplement Database. Source: SDM.

**AVELF:** Average of five different indices of ethnolinguistic fragmentation which is the probability of two random people in a country not coming from the same ethnolinguistic group; calculated by Easterly and Levine (1997). Source: SDM.

**CONFUC:** Fraction of Confucian population in 1970 and 1980. Calculated by Barro (1999) from World Christian Encyclopedia. Source: SDM.

**MINING:** Fraction of GDP produced in the Mining and Quarrying sector (including oil and gas; data are for the year 1988 when possible, or the closest available year); calculated by Hall and Jones (1999). Source: SDM.

**PRIGHTS:** Index of political rights comprising rights to vote, compete for public offices and for elected representatives to have a decisive vote on public policies (from 1 to 7; 1 = most rights). Calculated by Barro and Lee (1993) from GASTIL. Source: SDM.

**MALARIA:** Index of malaria prevalence in 1966, which is the product of the fraction of land area subject to malaria times the fraction of falciparum malaria cases. Calculated by Gallup et al. (2001) from World Health Organization data. Source: SDM.

All data used in this study can be downloaded from the project’s website mentioned at the end of Section 2.

## References

- Barro, R., 1991. Economic growth in a cross section of countries. *Quarterly Journal of Economics* 106, 407–433.
- Barro, R., 1999. Determinants of democracy. *Journal of Political Economy* 107, S158–S183.
- Barro, R., Lee, J., 1993. International comparisons of educational attainment. *Journal of Monetary Economics* 32, 363–394.
- Bauwens, L., Lubrano, M., Richard, J.-F., 1999. Bayesian Inference in

- Dynamic Econometric Models. Oxford University Press, Oxford.
- Brock, W., Durlauf, S.N., 2001. Growth empirics and reality. *World Bank Economic Review* 15, 229–272.
- Claeskens, G., Hjort, N.L., 2003. The focused information criterion. *Journal of the American Statistical Association* 98, 900–916.
- Crespo-Cuaresma, J., Doppelhofer, G., 2007. Nonlinearities in cross-country growth regressions: A Bayesian averaging of thresholds (BAT) approach. *Journal of Macroeconomics* 29, 541–554.
- 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., Magnus, J.R., 2004. On the harm that ignoring pretesting can cause. *Journal of Econometrics* 122, 27–46.
- De Long, J., Summers, L., 1991. Equipment investment and economic growth. *Quarterly Journal of Economics* 106, 445–502.
- Draper, D., 1995. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society (Series B)* 57, 45–97.
- Doppelhofer, G., Weeks, M., 2008. Robust model averaging. Paper presented at the CESifo Area Conference on Macro, Money and International Finance, 14–15 March 2008, Munich.
- Durlauf, S.N., Johnson, P.A., Temple, J.R.W., 2005. Growth econometrics. In: Aghion, P., Durlauf, S.N. (Eds.), *Handbook of Economic Growth*. North Holland, Amsterdam, pp. 555–677.
- Durlauf, S.N., Kourtellos, A., Tan, C.M., 2008a. Empirics of growth and development. In: Dutt, A.K., Ros, J. (Eds.), *International Handbook of Development Economics*, Vol. 1. Edward Elgar, Cheltenham, UK, forthcoming.
- Durlauf, S.N., Kourtellos, A., Tan, C.M., 2008b. Are any growth theories robust? *Economic Journal* 118, 329–346.
- Easterly, W., Levine, R., 1997. Africa’s growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics* 112, 1203–1250.
- Eicher, T., Henn, C., Papageorgiou, C., 2007a. Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects. UWEC 2007-18, University of Washington.

- Eicher, T., Papageorgiou, C., Raftery, A.E., 2007b. Determining growth determinants: Default priors and predictive performance in Bayesian model averaging. UWEC 2007-25, University of Washington.
- Eicher, T., Papageorgiou, C., Roehn, O., 2007c. Unraveling the fortunes of the fortunate: An iterative Bayesian model averaging (IBMA) approach. *Journal of Macroeconomics* 29, 494–514.
- Fernández, C., Ley, E., Steel, M.F.J., 2001a. Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16, 563–576.
- Fernández, C., Ley, E., Steel, M.F.J., 2001b. Benchmarks priors for Bayesian model averaging. *Journal of Econometrics* 100, 381–427.
- Gallup, J., Mellinger, A., Sachs, J., 2001. Geography Datasets. Center for International Development at Harvard University (CID). Data website: <http://www2.cid.harvard.edu/ciddata/geographydata.htm>.
- Hall, R.E., Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* 114, 83–116.
- Hansen, B.E., 2007. Least squares model averaging. *Econometrica* 75, 1175–1189.
- Heston, A., Summers, R., Aten, B., 2001. Penn World Table version 6.0. Center for International Comparisons at the University of Pennsylvania (CICUP). Data website: <http://pwt.econ.upenn.edu>.
- Hjort, N.L., Claeskens, G., 2003. Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial (with discussion). *Statistical Science* 14, 382–417.
- Knack, S., Keefer, P., 1995. Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and Politics* VII, 207–227.
- Kormendi, R., Meguire, P., 1985. Macroeconomic determinants of growth: Cross-country evidence. *Journal of Monetary Economics* 16, 141–163.

- Leamer, E.E., 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.
- Leamer, E.E., 1985. Sensitivity analyses would help. *American Economic Review* 75, 308–313.
- León-González, R., Montolio, D., 2004. Growth, convergence and public investment. A Bayesian model averaging approach. *Applied Economics* 36, 1925–1936.
- Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82, 942–963.
- Ley, E., Steel, M.F.J., 2007. Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics* 29, 476–493.
- Ley, E., Steel, M.F.J., 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, forthcoming.
- Liang, H., Zou, G., Wan, A.T.K., Zhang, X., 2008. On optimal least squares modeling averaging. University of Rochester. Mimeo.
- Magnus, J.R., 1999. The traditional pretest estimator. *Theory of Probability and Its Applications* 44, 293–308.
- Magnus, J.R., 2002. Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal* 5, 225–236.
- Magnus, J.R., Durbin, J., 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67, 639–643.
- Magnus, J.R., Wan, A.T.K., Zhang, X., 2009. WALS estimation with non-spherical disturbances and an application to the Hong Kong housing market. Tilburg University, Mimeo.
- Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics* 107, 407–437.
- Masanjala, W., Papageorgiou, C., 2007. Initial conditions, and post-war growth in sub-Saharan Africa. Mimeo.

- Masanjala, W., Papageorgiou, C., 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.
- O’Hagan, A., 1994. *Bayesian Inference*. Kendall’s Advanced Theory of Statistics, Vol. 2B. Edward Arnold, London.
- Pericchi, L.R., Smith, A.F.M., 1992. Exact and approximate posterior moments for a normal location parameter. *Journal of the Royal Statistical Society (Series B)* 54, 793–804.
- Prüfer, P., Tondl, G., 2008. The FDI-growth nexus in Latin America: The role of source countries and local conditions. Center DP 2008-61, Tilburg University.
- Raftery, A.E., 1993. Bayesian model selection in structural equation models. In: Bollen, K., Long, J. (Eds.), *Testing Structural Equation Models*. Sage, Newbury Park, CA, pp. 163–180.
- Raftery, A.E., 1995. Bayesian model selection in social research. *Sociological Methodology* 25, 111-163.
- Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179–191.
- Sala-i-Martin, X., 1997. I just ran two million regressions. *American Economic Review* 87, 178–183.
- Sala-i-Martin, X., Doppelhofer, G., Miller, R.I., 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review* 94, 813–835.
- Solow, R.M., 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70, 65–94.
- Tsangarides, C.G., Ghura, D., Leite, C.A., 2004. *Is growth enough? Macroeconomic policy and poverty reduction*. International Monetary Fund, Washington, D.C.
- Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with  $g$ -prior distributions. In: Goel, P.K., Zellner A. (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. North-Holland, Amsterdam, pp. 233–243.



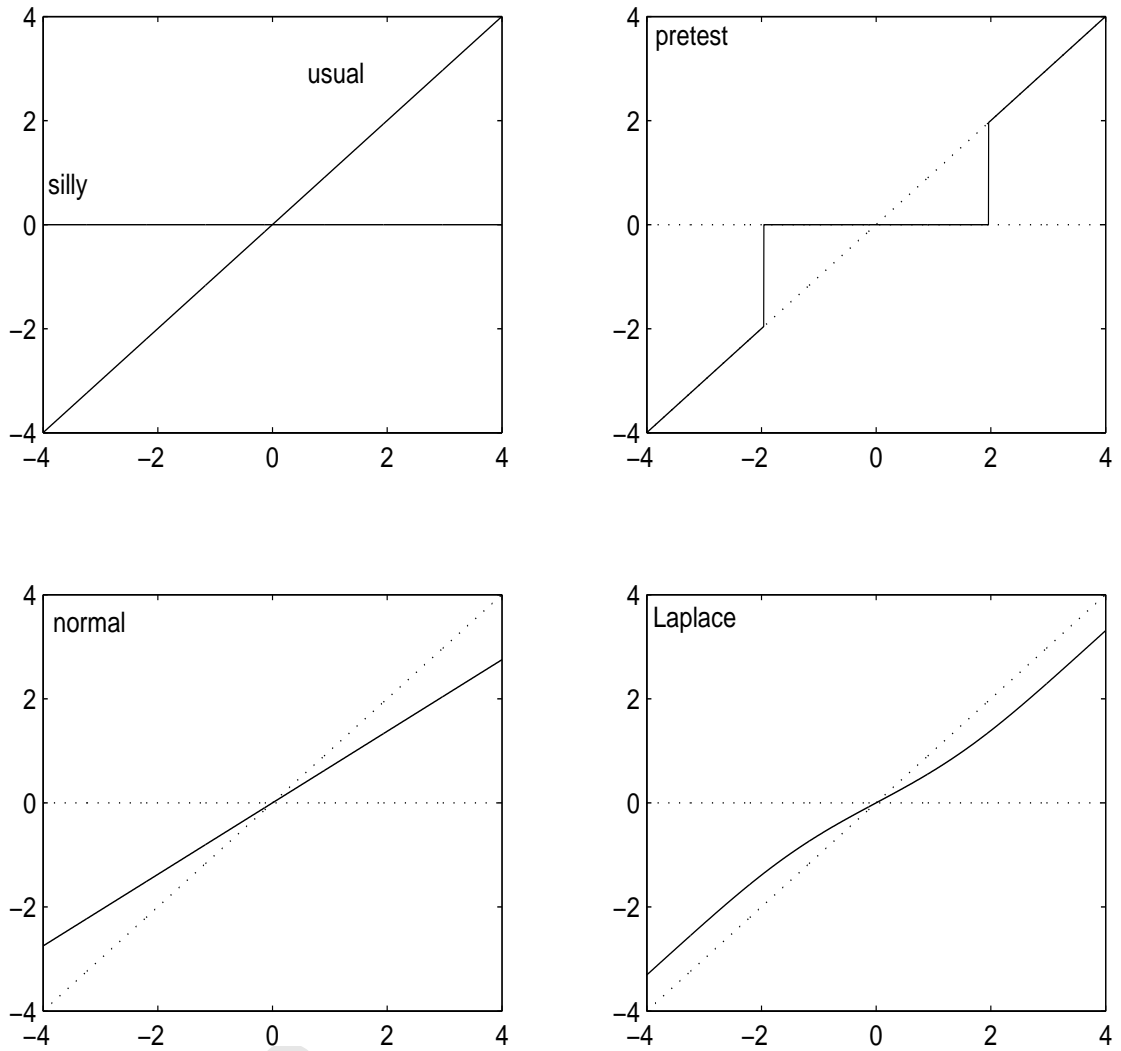
Figure 1: Five estimators  $t(x)$  of  $\eta$  when  $x \sim N(\eta, 1)$ .

Table 1: Model specifications, focus and auxiliary regressors.

Variable	Model 1	Model 2	Mean	SE
CONSTANT	Focus	Focus	1.0000	0.0000
GDP60	Focus	Focus	7.5253	0.8612
EQUIPINV	Focus	Focus	0.0432	0.0344
SCHOOL60	Focus	Focus	0.7807	0.2556
LIFE60	Focus	Focus	56.0676	1.1566
DPOP	Focus	Auxiliary	0.0206	0.0100
LAW	Auxiliary	Focus	0.5518	0.3332
TROPICS	Auxiliary	Focus	0.5481	0.4709
AVELF	Auxiliary	Focus	0.2984	0.2797
CONFUC	Auxiliary	Focus	0.0185	0.0862
MINING	—	Auxiliary	0.0482	0.0792
PRIGHTS	—	Auxiliary	3.4551	1.9073
MALARIA	—	Auxiliary	0.2866	0.4036

Table 2: Estimates  $\hat{\beta}$  and standard errors (in parentheses), Model 1, Set-up 1.

Regressor	Unrestricted	Restricted	GtS	WALS	BMA
<i>Focus regressors</i>					
CONSTANT	0.0609 (0.0223)	0.0587 (0.0242)	0.0518 (0.0214)	0.0594 (0.0221)	0.0492 (0.0229)
GDP60	-0.0155 (0.0033)	-0.0160 (0.0035)	-0.0145 (0.0032)	-0.0156 (0.0033)	-0.0139 (0.0035)
EQUIPINV	0.1366 (0.0552)	0.2405 (0.0583)	0.1377 (0.0555)	0.1555 (0.0551)	0.1644 (0.0615)
SCHOOL60	0.0170 (0.0098)	0.0184 (0.0111)	0.0191 (0.0097)	0.0175 (0.0097)	0.0160 (0.0102)
LIFE60	0.0008 (0.0004)	0.0010 (0.0004)	0.0008 (0.0004)	0.0009 (0.0004)	0.0008 (0.0004)
DPOP	0.3466 (0.2503)	-0.0341 (0.2611)	0.3275 (0.2513)	0.2651 (0.2487)	0.1654 (0.2770)
<i>Auxiliary regressors</i>					
LAW	0.0174 (0.0066)	—	0.0167 (0.0066)	0.0147 (0.0065)	0.0109 (0.0093)
TROPICS	-0.0075 (0.0040)	—	-0.0083 (0.0039)	-0.0055 (0.0037)	-0.0035 (0.0047)
AVELF	-0.0077 (0.0058)	—	—	-0.0053 (0.0048)	-0.0021 (0.0047)
CONFUC	0.0562 (0.0164)	—	0.0596 (0.0163)	0.0443 (0.0163)	0.0612 (0.0185)

Table 3: Estimates  $\hat{\beta}$  and standard errors (in parentheses), Model 1, Set-up 2.

Regressor	Unrestricted	Restricted	GtS	WALS	BMA
<i>Focus regressor</i>					
CONSTANT	0.0609 (0.0223)	0.0199 (0.0022)	0.0344 (0.0146)	0.0560 (0.0215)	0.0488 (0.0218)
<i>Auxiliary regressors</i>					
GDP60	-0.0155 (0.0033)	—	-0.0120 (0.0032)	-0.0136 (0.0033)	-0.0129 (0.0040)
EQUIPINV	0.1366 (0.0552)	—	0.1951 (0.0524)	0.1037 (0.0537)	0.1539 (0.0797)
SCHOOL60	0.0170 (0.0098)	—	—	0.0125 (0.0094)	0.0084 (0.0127)
LIFE60	0.0008 (0.0004)	—	0.0012 (0.0003)	0.0008 (0.0003)	0.0009 (0.0005)
DPOP	0.3466 (0.2503)	—	—	0.2236 (0.2156)	0.0261 (0.1252)
LAW	0.0174 (0.0066)	—	—	0.0137 (0.0063)	0.0090 (0.0092)
TROPICS	-0.0075 (0.0040)	—	—	-0.0055 (0.0039)	-0.0021 (0.0038)
AVELF	-0.0077 (0.0058)	—	—	-0.0083 (0.0057)	-0.0024 (0.0050)
CONFUC	0.0562 (0.0164)	—	0.0728 (0.0167)	0.0451 (0.0163)	0.0663 (0.0180)

Table 4: Estimates  $\hat{\beta}$  and standard errors (in parentheses), Model 2, Set-up 1.

Regressor	Unrestricted	Restricted	GtS	WALS	BMA
<i>Focus regressors</i>					
CONSTANT	0.0931 (0.0264)	0.0768 (0.0193)	0.0930 (0.0198)	0.0879 (0.0246)	0.0862 (0.0239)
GDP60	-0.0173 (0.0033)	-0.0156 (0.0033)	-0.0166 (0.0032)	-0.0167 (0.0033)	-0.0164 (0.0033)
EQUIPINV	0.1324 (0.0579)	0.1479 (0.0550)	0.1448 (0.0531)	0.1379 (0.0562)	0.1423 (0.0553)
CONFUC	0.0538 (0.0169)	0.0585 (0.0165)	0.0522 (0.0161)	0.0550 (0.0167)	0.0550 (0.0169)
SCHOOL60	0.0144 (0.0096)	0.0183 (0.0098)	0.0151 (0.0096)	0.0156 (0.0096)	0.0162 (0.0099)
LIFE60	0.0006 (0.0004)	0.0006 (0.0003)	0.0005 (0.0003)	0.0006 (0.0003)	0.0006 (0.0003)
LAW	0.0200 (0.0068)	0.0145 (0.0063)	0.0183 (0.0063)	0.0183 (0.0066)	0.0171 (0.0067)
TROPICS	-0.0055 (0.0041)	-0.0055 (0.0037)	-0.0029 (0.0037)	-0.0053 (0.0040)	-0.0044 (0.0041)
AVEFL	-0.0040 (0.0060)	-0.0073 (0.0059)	-0.0033 (0.0059)	-0.0049 (0.0059)	-0.0050 (0.0062)
<i>Auxiliary regressors</i>					
MINING	-0.0090 (0.0192)	---	---	-0.0056 (0.0149)	-0.0003 (0.0063)
DPOP	0.3352 (0.2542)	---	---	0.2147 (0.2178)	0.0650 (0.1705)
PRIGHTS	-0.0013 (0.0012)	---	---	-0.0008 (0.0010)	-0.0002 (0.0007)
MALARIA	-0.0104 (0.0052)	---	-0.0122 (0.0051)	-0.0075 (0.0050)	-0.0072 (0.0070)

Table 5: Estimates  $\hat{\beta}$  and standard errors (in parentheses), Model 2, Set-up 2.

Regressor	Unrestricted	Restricted	GtS	WALS	BMA
<i>Focus regressor</i>					
CONSTANT	0.0931 (0.0264)	0.0199 (0.0022)	0.0828 (0.0183)	0.0897 (0.0252)	0.0796 (0.0251)
<i>Auxiliary regressors</i>					
GDP60	-0.0173 (0.0033)	—	-0.0163 (0.0032)	-0.0156 (0.0033)	-0.0151 (0.0039)
EQUIPINV	0.1324 (0.0579)	—	0.1414 (0.0530)	0.0962 (0.0552)	0.1341 (0.0805)
CONFUC	0.0538 (0.0169)	—	0.0558 (0.0160)	0.0397 (0.0161)	0.0562 (0.0197)
SCHOOL60	0.0144 (0.0096)	—	—	0.0098 (0.0088)	0.0079 (0.0115)
LIFE60	0.0006 (0.0004)	—	0.0008 (0.0003)	0.0006 (0.0003)	0.0006 (0.0005)
LAW	0.0200 (0.0068)	—	0.0177 (0.0062)	0.0177 (0.0065)	0.0169 (0.0095)
TROPICS	-0.0055 (0.0041)	—	—	-0.0048 (0.0039)	-0.0006 (0.0021)
AVELF	-0.0040 (0.0060)	—	—	-0.0054 (0.0058)	-0.0006 (0.0027)
MINING	-0.0090 (0.0192)	—	—	-0.0063 (0.0182)	-0.0002 (0.0062)
DPOP	0.3352 (0.2542)	—	—	0.2146 (0.2175)	0.0179 (0.0990)
PRIGHTS	-0.0013 (0.0012)	—	—	-0.0012 (0.0010)	-0.0002 (0.0006)
MALARIA	-0.0104 (0.0052)	—	-0.0151 (0.0046)	-0.0090 (0.0048)	-0.0129 (0.0068)

Table 6: Comparison of *pip*-values and *t*-ratios, Model 1.

Regressor	Set-up 1			Set-up 2		
	WALS	BMA		WALS	BMA	
	<i>t</i>	<i>pip</i>	<i>t</i>	<i>t</i>	<i>pip</i>	<i>t</i>
CONSTANT	2.69	1.00	2.15	2.60	1.00	2.24
GDP60	-4.78	1.00	-3.96	-4.16	0.98	-3.21
EQUIPINV	2.82	1.00	2.67	1.93	0.88	1.93
SCHOOL60	1.80	1.00	1.58	1.33	0.14	0.66
LIFE60	2.44	1.00	2.32	2.34	0.40	1.82
DPOP	1.07	1.00	0.60	1.04	0.85	0.21
LAW	2.25	0.68	1.17	2.18	0.59	0.98
TROPICS	-1.49	0.45	-0.75	-1.41	0.32	-0.55
AVELF	-1.11	0.25	-0.44	-1.45	0.27	-0.48
CONFUC	2.72	0.99	3.31	2.77	0.99	3.69

Table 7: Comparison of BACE, BMA, and WALS, SDM data.

Regressor	BACE	BMA- $u$	BMA- $c$	WALS- $F1$	WALS- $F8$
01 East Asian dummy	0.0218 (0.0061)	0.0180 (0.0092)	0.0211 (0.0056)	0.0102 (0.0158)	0.0092 (0.0159)
02 Primary schooling 1960 ( $F$ )	0.0269 (0.0080)	0.0242 (0.0116)	0.0278 (0.0075)	0.0205 (0.0167)	0.0293 (0.0175)
03 Investment price ( $F$ )	-0.0001 (0.0000)	-0.0001 (0.0000)	-0.0001 (0.0000)	-0.0001 (0.0000)	-0.0001 (0.0000)
04 GDP 1960 (log) ( $F$ )	-0.0085 (0.0029)	-0.0065 (0.0041)	-0.0083 (0.0026)	-0.0067 (0.0075)	-0.0093 (0.0077)
05 Fraction of tropical area ( $F$ )	-0.0148 (0.0042)	-0.0101 (0.0079)	-0.0153 (0.0037)	-0.0097 (0.0156)	-0.0123 (0.0158)
06 Population density coastal 1960's	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
07 Malaria prevalence in 1960's	-0.0157 (0.0062)	-0.0033 (0.0070)	-0.0162 (0.0060)	0.0054 (0.0111)	0.0055 (0.0112)
08 Life expectancy in 1960 ( $F$ )	0.0008 (0.0004)	0.0001 (0.0003)	0.0008 (0.0003)	-0.0000 (0.0006)	0.0001 (0.0006)
09 Fraction Confucian ( $F$ )	0.0544 (0.0224)	0.0088 (0.0218)	0.0545 (0.0212)	0.0241 (0.0372)	0.0470 (0.0387)
10 African dummy	-0.0147 (0.0069)	-0.0023 (0.0060)	-0.0157 (0.0065)	-0.0059 (0.0130)	-0.0053 (0.0131)
11 Latin American dummy	-0.0128 (0.0058)	-0.0019 (0.0052)	-0.0139 (0.0051)	-0.0086 (0.0270)	-0.0084 (0.0270)
12 Fraction GDP in mining	0.0388 (0.0193)	0.0040 (0.0133)	0.0394 (0.0191)	0.0342 (0.0408)	0.0356 (0.0411)
13 Spanish colony	-0.0107 (0.0050)	-0.0012 (0.0037)	-0.0112 (0.0046)	0.0076 (0.0148)	0.0077 (0.0148)
14 Years open	0.0122 (0.0063)	0.0008 (0.0034)	0.0112 (0.0058)	-0.0020 (0.0145)	-0.0030 (0.0145)
15 Fraction Muslim	0.0126 (0.0063)	0.0010 (0.0039)	0.0125 (0.0060)	0.0073 (0.0132)	0.0077 (0.0132)
16 Fraction Buddhist	0.0217 (0.0107)	0.0018 (0.0065)	0.0208 (0.0100)	0.0157 (0.0258)	0.0149 (0.0256)
17 Ethnolinguistic fractionalization ( $F$ )	-0.0113 (0.0058)	-0.0011 (0.0037)	-0.0111 (0.0051)	-0.0005 (0.0112)	-0.0016 (0.0117)
18 Government consumption share 1960's	-0.0442 (0.0254)	-0.0031 (0.0124)	-0.0415 (0.0214)	0.5530 (0.7560)	0.5538 (0.7560)
19 Population density 1960	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
20 Real exchange rate distortions	-0.0001 (0.0000)	-0.0000 (0.0000)	-0.0001 (0.0000)	-0.0000 (0.0001)	-0.0000 (0.0001)
21 Fraction speaking foreign language	0.0070 (0.0040)	0.0004 (0.0018)	0.0067 (0.0037)	0.0048 (0.0078)	0.0046 (0.0078)
22 (Imports + exports)/GDP	0.0089 (0.0052)	0.0005 (0.0023)	0.0090 (0.0049)	0.0002 (0.0165)	0.0003 (0.0164)
23 Political rights	-0.0018 (0.0012)	-0.0001 (0.0006)	-0.0020 (0.0010)	-0.0007 (0.0026)	-0.0006 (0.0025)
24 Government share of GDP	-0.0349 (0.0294)	-0.0017 (0.0090)	-0.0364 (0.0228)	-0.5725 (0.7591)	-0.5720 (0.7591)
25 Higher education in 1960	-0.0697 (0.0418)	-0.0032 (0.0169)	-0.0711 (0.0383)	-0.0490 (0.0735)	-0.0491 (0.0734)
26 Fraction population in tropics	-0.0107 (0.0068)	-0.0003 (0.0020)	-0.0100 (0.0066)	-0.0055 (0.0151)	-0.0051 (0.0149)
27 Primary exports in 1970	-0.0113 (0.0075)	-0.0003 (0.0023)	-0.0113 (0.0074)	-0.0041 (0.0145)	-0.0042 (0.0143)
28 Public investment share	-0.0615 (0.0430)	-0.0029 (0.0161)	-0.0699 (0.0405)	-0.0746 (0.0578)	-0.0738 (0.0571)
29 Fraction Protestant	-0.0119 (0.0093)	-0.0003 (0.0021)	-0.0110 (0.0086)	0.0065 (0.0266)	0.0063 (0.0267)
30 Fraction Hindu	0.0176 (0.0126)	0.0004 (0.0030)	0.0163 (0.0117)	0.0061 (0.0276)	0.0059 (0.0273)
31 Fraction population less than 15	0.0450 (0.0411)	0.0012 (0.0094)	0.0462 (0.0345)	0.0578 (0.1177)	0.0579 (0.1177)
32 Air distance to big cities	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
33 Government consumption share deflated with GDP prices	-0.0336 (0.0274)	-0.0006 (0.0053)	-0.0305 (0.0245)	-0.0281 (0.0587)	-0.0291 (0.0590)
34 Absolute latitude	0.0001 (0.0002)	0.0000 (0.0000)	0.0001 (0.0002)	-0.0002 (0.0004)	-0.0002 (0.0004)
35 Fraction Catholic	-0.0084 (0.0085)	-0.0001 (0.0012)	-0.0070 (0.0082)	0.0003 (0.0230)	-0.0001 (0.0230)
36 Fertility in 1960's	-0.0075 (0.0101)	-0.0001 (0.0015)	-0.0074 (0.0093)	-0.0058 (0.0218)	-0.0057 (0.0219)



Comparison of BACE, BMA, and WALIS, SDM data—continued.

Regressor	BACE	BMA- <i>u</i>	BMA- <i>c</i>	WALIS- <i>F1</i>	WALIS- <i>F8</i>
37 European dummy	-0.0023 (0.0105)	-0.0000 (0.0011)	-0.0008 (0.0093)	0.0080 (0.0307)	0.0091 (0.0309)
38 Outward orientation	-0.0033 (0.0027)	-0.0001 (0.0005)	-0.0033 (0.0025)	-0.0005 (0.0057)	-0.0008 (0.0055)
39 Colony dummy	-0.0050 (0.0047)	-0.0001 (0.0007)	-0.0044 (0.0041)	-0.0039 (0.0113)	-0.0031 (0.0112)
40 Civil liberties	-0.0072 (0.0071)	-0.0001 (0.0013)	-0.0076 (0.0071)	-0.0031 (0.0128)	-0.0026 (0.0130)
41 Revolutions and coups	-0.0071 (0.0061)	-0.0001 (0.0011)	-0.0069 (0.0056)	-0.0107 (0.0128)	-0.0104 (0.0128)
42 British colony	0.0037 (0.0036)	0.0000 (0.0005)	0.0030 (0.0032)	0.0023 (0.0067)	0.0028 (0.0067)
43 Hydrocarbon deposits in 1993	0.0003 (0.0004)	0.0000 (0.0001)	0.0003 (0.0004)	0.0002 (0.0006)	0.0001 (0.0007)
44 Fraction population over 65	0.0194 (0.1195)	0.0002 (0.0110)	0.0156 (0.1062)	0.1114 (0.2363)	0.1115 (0.2364)
45 Defense spending share	0.0453 (0.0768)	0.0005 (0.0082)	0.0509 (0.0626)	0.5604 (0.7434)	0.5600 (0.7434)
46 Population in 1960	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
47 Terms of trade growth in 1960's	0.0326 (0.0467)	0.0004 (0.0058)	0.0358 (0.0447)	-0.0011 (0.0986)	-0.0015 (0.0983)
48 Public education spending/GDP in 1960's	0.1295 (0.1728)	0.0012 (0.0193)	0.1254 (0.1534)	0.5967 (0.7749)	0.5968 (0.7749)
49 Landlocked country dummy	-0.0021 (0.0042)	-0.0000 (0.0005)	-0.0029 (0.0036)	-0.0009 (0.0074)	-0.0006 (0.0073)
50 Religion measure	-0.0047 (0.0072)	-0.0001 (0.0009)	-0.0060 (0.0064)	0.0007 (0.0118)	0.0010 (0.0118)
51 Size of economy	-0.0005 (0.0014)	-0.0000 (0.0001)	-0.0004 (0.0013)	-0.0025 (0.0047)	-0.0021 (0.0047)
52 Socialist dummy	0.0040 (0.0050)	0.0000 (0.0006)	0.0039 (0.0047)	-0.0013 (0.0087)	-0.0011 (0.0087)
53 English-speaking population	-0.0037 (0.0071)	-0.0000 (0.0008)	-0.0028 (0.0073)	0.0041 (0.0106)	0.0048 (0.0105)
54 Average inflation 1960–1990	-0.0001 (0.0001)	-0.0000 (0.0000)	-0.0001 (0.0001)	0.0001 (0.0004)	0.0001 (0.0004)
55 Oil-producing country dummy	0.0048 (0.0071)	0.0000 (0.0007)	0.0040 (0.0058)	0.0018 (0.0141)	0.0010 (0.0140)
56 Population growth rate 1960–1990	0.0208 (0.3078)	0.0002 (0.0243)	0.0257 (0.2634)	-0.0664 (0.7732)	-0.0664 (0.7732)
57 Timing of independence	0.0011 (0.0021)	0.0000 (0.0002)	0.0011 (0.0018)	-0.0005 (0.0049)	-0.0007 (0.0048)
58 Fraction land area near navigable water	-0.0026 (0.0059)	-0.0000 (0.0005)	-0.0014 (0.0053)	-0.0030 (0.0135)	-0.0036 (0.0133)
59 Square of inflation 1960–1990	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
60 Fraction spent in war 1960–1990	-0.0014 (0.0092)	-0.0000 (0.0007)	-0.0008 (0.0083)	0.0015 (0.0195)	0.0021 (0.0194)
61 Land area	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
62 Tropical climate zone	-0.0021 (0.0066)	-0.0000 (0.0005)	-0.0007 (0.0058)	-0.0036 (0.0097)	-0.0039 (0.0097)
63 Terms of trade ranking	-0.0037 (0.0096)	-0.0000 (0.0008)	-0.0029 (0.0086)	-0.0030 (0.0219)	-0.0025 (0.0218)
64 Capitalism	-0.0002 (0.0011)	-0.0000 (0.0001)	-0.0005 (0.0010)	-0.0014 (0.0020)	-0.0016 (0.0019)
65 Fraction Orthodox	0.0057 (0.0136)	0.0000 (0.0011)	0.0049 (0.0121)	0.0020 (0.0190)	0.0022 (0.0187)
66 War participation 1960–1990	-0.0007 (0.0030)	-0.0000 (0.0002)	-0.0004 (0.0027)	-0.0009 (0.0050)	-0.0006 (0.0049)
67 Interior density	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)

Table 8: Comparison of absolute  $t$ -ratios, SDM data.

Regressor	BACE	BMA- $u$	BMA- $c$	WALS- $F1$	WALS- $F8$
02 SCHOOL60 ( $F$ )	3.37	2.08	3.71	1.23	1.67
03 IPRICE ( $F$ )	3.36	1.99	3.62	2.21	2.47
04 GDP60 ( $F$ )	2.96	1.58	3.17	0.90	1.21
05 TROPICS ( $F$ )	3.49	1.29	4.12	0.62	0.78
07 MALARIA	2.54	0.46	2.70	0.48	0.50
08 LIFE60 ( $F$ )	2.28	0.40	2.30	0.06	0.12
09 CONFUC ( $F$ )	2.43	0.40	2.57	0.65	1.22
12 MINING	2.02	0.30	2.06	0.84	0.87
17 AVELF ( $F$ )	1.93	0.30	2.15	0.04	0.14
23 PRIGHTS	1.54	0.24	1.94	0.27	0.25
56 DPOP	0.07	0.01	0.10	0.09	0.09