Tilburg University

Concept-Based Bayesian Model Averaging and Growth Empirics

Magnus, J.R.; Wang, W.

Publication date:
2012

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Magnus, J. R., & Wang, W. (2012). *Concept-Based Bayesian Model Averaging and Growth Empirics*. (CentER Discussion Paper; Vol. 2012-017). Econometrics.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2012-017

**CONCEPT-BASED BAYESIAN MODEL
AVERAGING AND GROWTH EMPIRICS**

By

Jan R. Magnus, Wendun Wang

February 2012

ISSN 0924-7815

Concept-based Bayesian model averaging and growth empirics*

February 14, 2012

Jan R. Magnus

*Department of Econometrics & Operations Research,
Tilburg University*

Wendun Wang

CentER, Tilburg University

*E-mail addresses: magnus@uvt.nl (Jan Magnus), wangwendun@gmail.com (Wendun Wang).

Corresponding author:

Jan R. Magnus
Department of Econometrics & Operations Research
Tilburg University
PO Box 90153
5000 LE Tilburg
The Netherlands
E-mail: magnus@uvt.nl

Abstract: In specifying a regression equation, we need to determine which regressors to include, but also how these regressors are measured. This gives rise to two levels of uncertainty: concepts (level 1) and measurements within each concept (level 2). In this paper we propose a hierarchical weighted least squares (HWALS) method to address these uncertainties. We examine the effects of different growth theories taking into account the measurement problem in the growth regression. We find that estimates produced by HWALS provide intuitive and robust explanations. We also consider approximation techniques when the number of variables is large or when computing time is limited, and we propose possible strategies for sensitivity analysis.

JEL Classification: C51, C52, C13, C11.

Keywords: Hierarchical model averaging, Growth determinants, Measurement problem.

1 Introduction

In applied econometrics, when one specifies the equation to be estimated, one has to decide which *concepts* (say inflation or education) to include in the regression; and one has to decide which *measurements* of these concepts to use. In practice, one would often try many different concepts, and select the most appealing combination. This method is called ‘pretesting’ and it is a dangerous method to apply, because one reports estimates conditional on the selected model, but interprets them as unconditional estimates. Bayesian model averaging allows us to combine model selection and estimation into *one* procedure, thus avoiding the pretest problem. But, as we have just seen, we have, in addition to the pretest problem (which concepts to use?), also a ‘measurement problem’ (which measurement to use for each concept?).

This paper addresses both problems by introducing hierarchical (two-level) model averaging, where we perform model averaging over concepts *and* measurements. From here on we shall denote concepts as *groups*, and measurements as *variables*. We propose hierarchical weighted least squares (HWALS), a generalization of WALS (Magnus et al., 2010). In hierarchical model averaging, we introduce prior probabilities for the variables in each group, and treat the regression parameters as hierarchical random variables. We are uncertain about the error term, about which groups to select, and about which variables to select. All three levels of uncertainty are explicitly taken into account in hierarchical WALS estimation.

The HWALS procedure has several advantages. It provides an estimate and standard deviation for each group, which facilitates statistical inference and enables us to analyze the effect of each group; it combines model selection and estimation and thus avoids the problem of pretesting (Danilov and Magnus, 2004); it provides a clear economic explanation of the estimates, which possess well-defined finite-sample properties, unlike some other related methods (factor analysis, meta-analysis); and its computational burden is very light, especially compared to standard Bayesian model averaging and Bayesian averaging of classical estimates (BACE).

In the empirical growth literature the three types of uncertainty are especially important. The uncertainty on the choice of group occurs because growth theories are open-ended (Brock and Durlauf, 2001), while the uncertainty on the choice of variable occurs because different researchers employ different empirical proxies to measure the same group. As a result, there is currently no consensus on the question which determinants are most salient

and robust in explaining cross-country differences (Durlauf et al., 2005). In addition, the number of variables in growth empirics is large and may even exceed the number of observations. For example, Durlauf et al. (2005) list 145 regressors, while the number of countries is typically less in cross-country growth studies. Our paper investigates the effects of various growth determinants, explicitly taking into account the three types of uncertainty. A recent study by Durlauf et al. (2008) has a similar purpose, but their approach is different in that it does not distinguish between groups and variables, and hence does not produce one estimate and standard deviation for each group.

We mainly compare our estimates with those of Sala-i-Martin et al. (2004), hereafter SDM, and with the WALS estimates of Magnus et al. (2010). Our empirical results are different but generally not in conflict with this literature. However, our hierarchical model averaging estimates produce more intuitive signs and they are more robust. This is the benefit we gain from not ignoring the measurement problem, so that correlated variables within one group are not all included in the regression. We find evidence to support the relative robustness and importance of proximate determinants, such as health and initial state. We also find that some regional effects are strong, such as the East Asian and the colony dummy. These findings are consistent with most empirical growth theories. In contrast to the current literature we find that education and government intervention are not robust, because some of the variables in these groups have poor explanatory power in the growth regression.

The paper is organized as follows. In Section 2 we present the hierarchical estimation strategy. Section 3 describes the data, grouping, and scaling. We apply our estimation strategy to the data in Section 4 and discuss the results. Next we address the potential problem that the number of variables is too large to apply the HWALS technique directly. In that case, approximations are required and these are discussed in Section 5. A key issue is how sensitive the results are with respect to the assumptions made. We investigate the sensitivity with respect to prior probabilities (Section 6) and with respect to grouping (Section 7). Section 8 concludes. Our background document (Magnus and Wang, 2012) provides further and more detailed information.

2 Hierarchical weighted average least squares

2.1 Groups and variables

We write the linear regression model as

$$y = X^* \beta^* + \epsilon = X_1^* \beta_1^* + X_2^* \beta_2^* + \epsilon, \quad (1)$$

where we note two deviations from standard notation. First we write X^* and β^* rather than X and β , because the regressors are considered to be ‘groups’, for example education or inflation. These are groups (concepts) rather than precisely defined variables. There are many measures of education and of inflation that the researcher could use. These measurements of the same concept in one group are our ‘variables’. Second, we distinguish between focus regressors (labeled 1) and auxiliary regressors (labeled 2). Focus regressors are in the model irrespective of any preliminary test or diagnostic. These include the variables of specific interest and the variables that economic knowledge dictates to be in the model. Auxiliary regressors, on the other hand, may or may not be in the model, depending on prior knowledge and diagnostics.

We write the columns of the (group) regressors as

$$X_1^* = (x_{1,1}^*, \dots, x_{1,k_1}^*), \quad X_2^* = (x_{2,1}^*, \dots, x_{2,k_2}^*), \quad (2)$$

and the components of the (group) parameter vectors as

$$\beta_1^* = \begin{pmatrix} \beta_{1,1}^* \\ \beta_{1,2}^* \\ \vdots \\ \beta_{1,k_1}^* \end{pmatrix}, \quad \beta_2^* = \begin{pmatrix} \beta_{2,1}^* \\ \beta_{2,2}^* \\ \vdots \\ \beta_{2,k_2}^* \end{pmatrix}. \quad (3)$$

Important is the distinction between *groups* and *variables*. The l_1 -th focus group x_{1,l_1}^* contains m_{1,l_1} variables, and the l_2 -th auxiliary group x_{2,l_2}^* contains m_{2,l_2} variables. Groups may contain only one variable, but many groups will contain several variables. While the variables themselves are considered deterministic, a group is random (if there are at least two variables in the group) because the choice between the variables or the weighting scheme depends on the data (and on priors).

We attach prior probabilities to the variables based on our confidence. Thus,

$$\Pr(x_{1,l_1}^* = x_{1,l_1}^i) = \pi_{1,l_1}^i, \quad \Pr(x_{2,l_2}^* = x_{2,l_2}^j) = \pi_{2,l_2}^j, \quad (4)$$

where $i = 1, \dots, m_{1,l_1}$ and $j = 1, \dots, m_{2,l_2}$, under the constraints

$$\sum_{i=1}^{m_{1,l_1}} \pi_{1,l_1}^i = 1, \quad \sum_{j=1}^{m_{2,l_2}} \pi_{2,l_2}^j = 1. \quad (5)$$

When only one variable is available in a group, then there is no ambiguity about the choice of variable, and we set $m_{1,l_1} = 1$ or $m_{2,l_2} = 1$ and correspondingly $\pi_{1,l_1}^i = 1$ or $\pi_{2,l_2}^j = 1$.

Given specific variables x_{1,l_1}^i and x_{2,l_2}^j in each group, we construct the design matrices

$$X_1^{(i)} = \left(x_{1,1}^{i_1}, \dots, x_{1,k_1}^{i_{k_1}} \right), \quad X_2^{(j)} = \left(x_{2,1}^{j_1}, \dots, x_{2,k_2}^{j_{k_2}} \right), \quad (6)$$

and the parameter vectors

$$\beta_1^{(i)} = \begin{pmatrix} \beta_{1,1}^{i_1} \\ \beta_{1,2}^{i_2} \\ \vdots \\ \beta_{1,k_1}^{i_{k_1}} \end{pmatrix}, \quad \beta_2^{(j)} = \begin{pmatrix} \beta_{2,1}^{j_1} \\ \beta_{2,2}^{j_2} \\ \vdots \\ \beta_{2,k_2}^{j_{k_2}} \end{pmatrix}, \quad (7)$$

where $(i) = (i_1, \dots, i_{k_1})$ and $(j) = (j_1, \dots, j_{k_2})$. The resulting model can then be written as

$$y = X_1^{(i)} \beta_1^{(i)} + X_2^{(j)} \beta_2^{(j)} + \epsilon, \quad (8)$$

where we emphasize that each model includes precisely *one* variable from each group, thus avoiding problems caused by including many highly correlated variables.

2.2 A three-step procedure

Under the assumption that the prior distributions on separate groups are independent, the prior probability attached to a specific choice of variables (i, j) is given by

$$\pi^{(i,j)} = \prod_{l_1=1}^{k_1} \pi_{1,l_1}^{i_{l_1}} \prod_{l_2=1}^{k_2} \pi_{2,l_2}^{j_{l_2}}. \quad (9)$$

This is the first step. For given (i) and (j) we estimate (8) by Bayesian model averaging. In Bayesian model averaging the estimates are computed as weighted averages of the estimates obtained over all possible models, thus allowing for the fact that auxiliary regressors may or may not be in the model, depending on priors and diagnostics. One major advantage of Bayesian model averaging is that it treats model selection and estimation as *one* procedure. We shall use a method called WALS (weighted average least squares), but this is not essential in the development. The advantage of WALS is both conceptual and computational. The version of WALS employed here is described in Magnus et al. (2010), and the estimates are made scale-independent using the weighting scheme proposed in De Luca and Magnus (2011). Thus we obtain the posterior mean (the WALS estimates),

$$b^{(i,j)} = \begin{pmatrix} \hat{\beta}_1^{(i,j)} \\ \hat{\beta}_2^{(i,j)} \end{pmatrix}, \quad (10)$$

and the posterior variance matrix $V^{(i,j)}$. That is the second step. These posterior moments are, of course, still conditional on the choice of variables, that is, on (i, j) . In the third and final step we obtain the unconditional posterior moments b and V from

$$b = \sum_{(i,j)} \pi^{(i,j)} b^{(i,j)} \quad (11)$$

and

$$V = \sum_{(i,j)} \pi^{(i,j)} \left(V^{(i,j)} + b^{(i,j)} b^{(i,j)'} \right) - b b'. \quad (12)$$

The variance V in the posterior distribution thus fully represents the three sources of uncertainty associated with the hierarchical procedure: uncertainty represented by the error term, given the specification of the model; uncertainty about which auxiliary groups to include; and uncertainty about which variables to include in each group. The estimator b is the hierarchical WALs (HWALS) estimator, and V is taken to be its variance.

2.3 Choice of π

The prior probabilities π should be specified, and the question is how. This depends on the strength of the researcher's prior information and belief. We distinguish between four cases. In the first case, the researcher has no prior information at all. One measurement of the group variable is as likely as the other, so we assign equal weights within each group, that is,

$$\pi_{1,l_1}^i = \frac{1}{m_{1,l_1}}, \quad \pi_{2,l_2}^j = \frac{1}{m_{2,l_2}}. \quad (13)$$

This is our default. The second case occurs when we have unequal prior information about the variables, in particular there may be one or two variables that are known a priori to be particularly important. In the third case we can rank the prior probabilities within one group, but have no knowledge of their precise values.

In the fourth case we propose to use data-dependent priors. We write $X_1^* = (X_{11} : X_{12}^*)$, where X_{11} contains the focus regressors for which only one variable is available, and X_{12}^* contains the focus regressors for which at least two variables are available. For each group l in X_{12}^* we estimate

$$y = X_{11} \beta_{11} + \beta_{1,l}^i x_{1,l}^i + \epsilon \quad (i = 1, \dots, m_{1,l}), \quad (14)$$

from which we calculate the likelihood $L(x_{1,l}^i) = \Pr(y, X | x_{1,l}^* = x_{1,l}^i)$. Then we update the prior $\pi_{1,l}^i$ by Bayes' rule:

$$\bar{\pi}_{1,l}^i = \Pr(x_{1,l}^* = x_{1,l}^i | y, X) = \frac{\pi_{1,l}^i L(x_{1,l}^i)}{\sum_{h=1}^{m_{1,l}} \pi_{1,l}^h L(x_{1,l}^h)}.$$

A larger weight is thus assigned to the variable with more explanatory power (larger likelihood).

We distinguish between two subcases. In case 4(a) (one-step updating) we update the priors for the auxiliary variables in the same way, based on the equation

$$y = X_{11}\beta_{11} + \beta_{2,l}^j x_{2,l}^j + \epsilon \quad (j = 1, \dots, m_{2,l}). \quad (15)$$

In case 4(b) (two-step updating) we update the priors for the auxiliary variables based on the extended equation

$$y = X_1^{(i)}\beta_1^{(i)} + \beta_{2,l}^j x_{2,l}^j + \epsilon \quad (j = 1, \dots, m_{2,l}), \quad (16)$$

where *all* focus groups are used, not only the groups with one variable (X_{11}), but also the groups with two or more variables. For the latter we select the variable with the highest posterior probability $\bar{\pi}_{1,l}^i$. We shall discuss the sensitivity of our results with respect to different choices of prior in Section 6.

3 Data, grouping, and scaling

We apply the proposed hierarchical method of Section 2 to growth empirics. There is a large literature on explaining cross-country growth differences, but this literature has not led to a consensus on which regressors to include, and, even if there is agreement on a regressor (group), there is still disagreement on which measurement (variable) of that regressor to use. These issues are well exposed in Brock and Durlauf (2001). Growth empirics thus provides a typical and important example of a situation where two types of uncertainty exist: uncertainty about the relevance of a group and uncertainty about which variable to select within the group.

Our data are taken primarily from SDM. The dependent variable is the average growth rate of GDP per capita 1960–1996. The SDM data set contains 88 countries and 67 variables (plus the constant term). To this list we have added seven variables from Sala-i-Martin (1997): six variables in education and one variable in government intervention. These are indicated with an asterisk (*) in Table 1. This makes a total of 74 variables (25 groups)

plus the constant term. We use 72 (rather than 88) countries, the maximum possible number if we wish to obtain a ‘balanced’ data set with an equal number of observations for all regressors. Since we have more variables than observations we can not estimate the whole set. Grouping will therefore be especially helpful here.

TABLES 1 and 2

The regressors are listed and grouped in Tables 1 and 2. The 74 variables are organized in 25 groups. The grouping is based on Durlauf et al. (2005) with two deviations: we split the ‘geography’ group in two (‘tropics effect’ and ‘geography excluding tropics effect’), and also the ‘government’ group (‘government intervention’ and ‘defense’). The reason for splitting up geography and government is that within the new groups ‘tropics effect’ and ‘government intervention’ the same concept is measured, while the remaining items are of a different nature.

We distinguish between two types of groups. A group of type I (Table 1) contains variables providing alternative measurements of one concept. For example, a country’s democracy (the concept) can be measured in several ways, and we allow two measurements (political rights and civil liberties). Another important growth determinant is education (the concept), which attempts to capture human capital accumulation. Since the output of human capital investment is difficult to measure, one typically resorts to input variables, such as the enrollment rate, school years, and the share of public education spending. These input variables serve as different measurements for the same concept. Measurements of the same concept are likely to be highly correlated, so we do not want to include all of them in our regression. In fact, we only want to use *one* measurement, but we do not know which one. Our theory of Section 2 applies to this type, that is to groups (1)–(12).

In contrast, a group of type II (Table 2) contains variables measuring different aspects of one concept. For example, the group ‘regional effect’ contains seven dummy variables, each indicating whether a country belongs to some particular (colonial) region. These variables all measure a regional effect, but a different aspect of it, and these aspects are not highly correlated or easily aggregated. Another example is the group ‘geography excluding tropics effect’ which contains three variables: fraction land area near navigable water, landlocked country dummy, and air distance to big cities, measuring different aspects of the region with low correlation. Our hierarchical theory does not apply to groups (13)–(20), because parameter estimates associated with these variables have different meanings, and hence a weighted sum of these estimates makes little sense. In our hierarchical estimation procedure

we can either include all variables of a type II group or select a representative. For groups (13)–(15) we choose one representative, namely ethnolinguistic fractionalization in the group ‘ethnicity and language’; religious intensity in the group ‘religion’ (Barro, 1999); and the openness measure in the group ‘trade statistics’. For groups (16)–(20) each variable captures a different aspect of the concept, and we include all variables. Groups (21)–(25) only contain one variable, and hence there is no difference between variable and group. In summary, we have 12 type I groups (35 variables) and 13 type II groups (39 variables).

Grouping of variables can be ambiguous. For example, one may argue that the share of public education spending in GDP (variable 8) should not be in the group ‘education’ but in the group ‘government intervention’, because it measures the government’s commitment to education, which is different from other measurements of education. Indeed, the correlation between the share of public education and other educational variables is small. We address such problems in Section 7, where we investigate *inter alia* the sensitivity of the posterior mean and variance to changes in the grouping.

Before we apply the hierarchical WALS procedure, we scale all variables, that is, we scale (and center) each variable x by replacing it with $(x - \text{mean}(x)) / \text{std}(x)$, so that the resulting transformed variable has zero mean and unit variance. In standard (non-hierarchical) WALS the centering has no effect (other than on the constant term), but the scaling does have an effect. This effect, however, can be removed by scaling the matrix

$$Z^{(i,j)} = X_2^{(j)'} \left(I - X_1^{(i)} \left(X_1^{(i)'} X_1^{(i)} \right)^{-1} X_1^{(i)'} \right) X_2^{(j)},$$

such that all its diagonal elements equal one (De Luca and Magnus, 2011). In hierarchical WALS the preliminary scaling is important because it makes the magnitudes of the estimated parameters within one group comparable.

TABLE 3

In addition to scaling the variables, we may also wish to change the sign of some variables, so that variables within one group are positively correlated. The five variables that have been re-signed are listed in Table 3. For example, in the group ‘health’ we change the definition of malaria prevalence to malaria non-prevalence, so that both variables in this group now measure the same thing rather than opposite things. When the group consists of more than two variables, this may not always be possible, but in our case it is. An example is the tropics effect, where we have three variables. A large value of absolute latitude implies a relatively small area in the tropics. But this is the

opposite of fraction of tropical area and tropical climate zone, the other two measurements in this group. Hence, we redefine absolute latitude so that its sign changes.

4 Growth empirics

There is not much consensus in the empirical growth literature on which growth determinants are salient and robust among a large set of growth theories. Most papers report insignificant coefficients for most determinants. One reason is that growth theories are open-ended (Durlauf et al., 2005), another that the same concept can be measured by (sometimes many) different empirical proxies. In this paper we concentrate on the second aspect. Different choices of measurement may result in very different estimates. If we include all or many measurements of the same concept in one regression, then the t -ratios will be misleading due to multicollinearity. Our theory allows us to treat the 74 (plus the constant) different measurements (variables) as elements of only 25 (plus the constant) concepts (groups).

We have to choose which groups are focus and which auxiliary, and we shall discuss two variants. In variant HWALS-F1 only the constant term is a focus group, while all other groups are auxiliary. This is the typical model averaging framework in which all explanatory variables are allowed to be either included or excluded, as in SDM (2004) or Ley and Steel (2007). More information is used in the second variant, HWALS-F8, where eight groups (including the constant term) are treated as focus groups. These groups are therefore in every model considered. The eight focus groups consist of four type I groups (education, health, initial state, tropics effect), two type II groups with a representative variable (ethnicity and language, religion), and two type II groups with only one variable (price distortion, constant term). Note that the distinction between focus and auxiliary is made at the group level. If a group is considered to be focus (auxiliary), then each variable in that group is also focus (auxiliary).

TABLES 4 and 5

In Tables 4 and 5 we present the results for HWALS-F1 and HWALS-F8, and compare them with WALS-F8. The WALS-F8 estimates are based on the 67 variables in SDM (2004), hence without the seven additional variables from Sala-i-Martin (1997). They differ from those in Magnus et al. (2010, Table 7), because of the scaling and the different number of observations. The WALS-F8 estimates correspond to variables; the estimates in

HWALS-F1 and HWALS-F8 to groups.

We shall also compare our results with the BACE estimates of SDM (2004). Since the posterior moments given by BACE are conditional on inclusion, their precision is misleading as pointed out in Magnus et al. (2010). Therefore, we compare with the *unconditional* BACE moments according to Equations (8) and (14) in SDM (2004). (The full set of unconditional BACE estimates is available in our background document.)

4.1 Sign comparisons

Let us first compare the signs of the estimates. The signs of the group estimates in HWALS-F1 and HWALS-F8 are highly correlated, except for the group ‘war’, but with a t -ratio of only 0.06 not much significance can be attached to this result. Next we compare the signs of HWALS-F8 with those of WALS-F8 and BACE, where we recall that the latter estimates are based on variables while the former are based on groups. We shall say that an HWALS estimate is ‘totally different’ from the BACE/WALS estimate if the sign of a type I group is opposite to *all* of its variables, and ‘partially different’ if the sign of a type I group is opposite to *some* of its variables. For type II groups this distinction is not necessary.

Comparing HWALS to WALS we see that in three of the 25 groups the estimates are partially different, and in no less than eleven groups they are totally different. Hence, quite different estimation results are produced by HWALS as compared to WALS. The signs produced by HWALS are generally more intuitive than those produced by WALS, except for education. For example, HWALS associates more tropical regions with a lower growth rate, while all variables of the tropics effect have a positive sign in WALS; HWALS finds that being more open has a positive impact on growth, while all variables in the trade policy indices have a negative sign in WALS; and HWALS finds that African and Latin-American countries generally grow slower and British colonial countries grow faster, while WALS reports the opposite.

For the group education, HWALS produces a negative (but not significant) estimate. This seems counterintuitive. Upon closer inspection we see that the education group contains many variables which are not robust and have relatively large standard deviations. We have nine education variables, and they measure education in three ways: the enrollment rate at different school levels (variables 5–7); educational attainment at different school levels (variables 9–13); and public spending on education (variable 8). Only the primary schooling enrollment rate in 1960 and the secondary school years have robust positive effects, while the signs of the remaining variables vary with the model specification. This is in line with most empirical growth literature,

although some care needs to be taken in explaining the strongly positive estimate of primary schooling in 1960 (Barro and Lee, 1993). The variation between different measurements and the insignificance of most measurements lead to an insignificant estimate of the group education. Therefore, the education effect on growth does not appear to be as robust as most studies suggest.

Comparing HWALS to BACE, we find three type I groups that are partially different (education, democracy, trade policy indices), and nine variables of type II groups that are totally different. Again, the signs produced by HWALS are more intuitive. For example, in contrast to BACE, HWALS finds a negative effect (not significant) of democracy (recall that civil liberties is re-signed, so that a positive estimate of democracy implies a negative effect), which supports recent studies (Barro, 1996); also in contrast to BACE, HWALS finds a positive correlation between growth and the European dummy; HWALS finds that more fraction GDP in mining leads to a lower growth rate, which is supported by most cross-country studies on the ‘curse of resource’, whereas BACE finds the opposite result; countries with more land area near navigable water have access to more convenient transportation and are typically more open, thus enhancing growth, as shown by HWALS but not by BACE; and HWALS finds a negative (but not significant) effect of defense spending share on economic growth, in line with recent studies in the military spending literature (Dunne and Uye, 2009).

4.2 Precision comparisons

Next we compare the t -ratios produced by HWALS-F8, WALs, and BACE (unconditional moments). The WALs and BACE t -ratios are largely similar. HWALS is generally more precise than WALs and BACE, especially for those groups/variables that are typically thought of as robust determinants.

For the focus groups, HWALS reports $t = 1.26$ for health, while the t -ratios of the two health variables (life expectancy and malaria prevalence) are 0.53 and -0.48 in WALs; and 0.45 and -0.53 in BACE. In the group ‘tropics effect’, the t -ratios of its three variables vary greatly in both WALs and BACE. Only the fraction of tropical area has a t -ratio slightly larger than 1 (in absolute value) in BACE, while the other two measurements all have $|t| < 0.30$. WALs even reports a counterintuitive positive effect. In contrast, HWALS combines three variables and gives a t -ratio of this group of approximately 1. The estimate of ‘ethnolinguistic fraction’ produced by HWALS has $|t| = 1.07$, while WALs and BACE show $|t| = 0.22$ and $|t| = 0.30$, respectively.

For the auxiliary groups, most estimates of type I groups produced by

HWALS are more significant than WALS and BACE (for example, demographic characteristics, inflation, scale effect). The estimates of all type II groups produced by HWALS are more significant than WALS and BACE.

4.3 Explanatory power

Particularly relevant is the contribution of various growth theories in explaining differences in cross-country growth rates. Since all variables are converted to the same scale, the estimates capture the explanatory power of each theory.

We find that investment price ($b = -0.0041$, $|t| = 2.4$) and the East Asian dummy ($b = 0.0058$, $|t| = 2.1$) are the most robust variables and explain most of the cross-country variation. Less robust but still strong in explanatory power are health ($b = 0.0073$, $|t| = 1.3$), initial state ($b = 0.0045$, $|t| = 0.7$), and the colony dummy ($b = -0.0038$, $|t| = 1.1$). These results provide evidence in favor of the neoclassical growth determinants, and they are also largely consistent with the findings in the conditional convergence literature and other related studies (Fernandez et al., 2001; Sala-i-Martin et al., 2004; Durlauf et al., 2008).

The groups tropics effect, ethnicity and language, African dummy, and terms of trade have slightly less explanatory power. Here our results differ from those in SDM, employing posterior inclusion probabilities, who report the first three variables to be among the most salient, and terms of trade to be among the least salient. The variables that are weakly or not related with growth are largely in line with the literature. Unlike most studies, education and government intervention are not robustly related to economic growth. In summary, the HWALS methodology yields intuitive and stable estimates.

5 Approximations for large k

To compute the HWALS estimates we need many runs of the WALS algorithm. Each run of the WALS algorithm requires model averaging over $k_2 = 41$ (HWALS-F1) or $k_2 = 34$ (HWALS-F8) auxiliary variables. In the case of BMA this would take much computing time (of the order 2^{k_2}), but in WALS much less (of the order k_2). This is one (but not the only one) advantage of WALS over BMA. Even so, in our application of the HWALS procedure, we have to repeat this algorithm $2^9 \times 3 \times 5 \times 9 = 69120$ times. This would be impossible with BMA or BACE, but it is still feasible in WALS, and the estimates reported in Tables 4 and 5 are based on exact computations.

If the number of groups and variables increases further, then estimating

all combinations becomes computationally too time-consuming, especially if we also want to perform simulations and sensitivity analyses. In such cases we have to resort to approximations. In this section we propose and compare several approximating algorithms. There are two aspects to the approximation: selecting the subset of WALS regressions to be performed and obtaining the corresponding WALS estimates; and assigning estimates to the non-sampled regressions based on the estimates of the sampled regressions. We shall discuss each aspect in turn.

5.1 Subset selection

Two types of subset selection are considered: non-probability sampling and probability sampling. The non-probability method chooses the combinations deterministically. We sample those combinations whose prior probabilities (weights) are larger than a predetermined critical value π^* , because these are the combinations composed of relatively ‘important’ variables in each group. We obtain WALS estimates for these combinations. The ‘precision’ of the approximation is controlled by

$$\alpha = \sum_{\pi^{(i,j)} > \pi^*} \pi^{(i,j)},$$

that is, the proportion of exact estimates in the approximated HWALS computation. Note that the non-probability approach does not work if we use equal weights within each group, because many combinations will then have exactly the same prior probability. We use two stopping rules. First, we reduce π^* until the precision α satisfies a required level α^* . Second, to bound computation time, we restrict the number of samples S by an upper bound S^* . Hence, we require $\alpha > \alpha^*$ and $S < S^*$.

In contrast, the probability method uses the weights as prior probabilities, and draws randomly (without replacement) from these weights. Each combination, also one with a low weight, can now be selected, but combinations with a high weight will have a higher selection probability than combinations with a low weight. The only requirement is $S < S^*$.

5.2 Approximating the non-sampled estimates

We consider two methods to approximate the non-sampled estimates from the sampled ones, first using neighboring estimates, then using a normalization of the probability. The first method is based on ‘neighboring’ estimates. For a given combination C , its ‘neighbors’ consist of those combinations containing

at least one group represented by a variable that is also present in C . The approximation averages the neighboring estimates.

Figure 1: Approximation using neighboring estimates

As an example we consider three groups G_1 , G_2 , and G_3 , containing 5, 2, and 3 variables, respectively, as in Figure 1. The combinations form a cube with 30 cells. Suppose we have sampled three combinations with corresponding estimates E_1 , E_2 , and E_3 , and that P is the estimate to be approximated based on E_1 , E_2 and E_3 . We compute P in two steps, first averaging within each dimension that contains sampled estimates, and then averaging over all dimensions. In our case,

$$P = \frac{(E_1 + E_2)/2 + E_3}{2}. \quad (17)$$

This averaging scheme is better than a one-step average $(E_1 + E_2 + E_3)/3$, because P puts equal weight on each dimension, while the one-step average puts more weight on the dimension containing more sampled estimates. (A comparison between these two averaging methods is given in our background document.) Neighboring estimates are good approximates because changing the measurement of a group has a much smaller impact on estimates of other groups (indirect effect) than it does on the group itself (direct effect); see also Section 6 below. This method works unless an estimate to be approximated does not have any neighbors that are sampled. Such estimates are approximated by the average over all sampled estimates.

In the second method we normalize the probability of the sampled combinations, so that the sum of these probabilities equals 1, that is,

$$\pi_*^{(i,j)} = \frac{\pi^{(i,j)}}{\sum_{(m,n) \in \mathcal{C}} \pi^{(m,n)}}, \quad (i,j) \in \mathcal{C}, \quad (18)$$

where $\mathcal{C}$ is the set of sampled combinations. This implies that all approximated estimates are equal. In particular, suppose we have S samples from T combinations, and denote the sampled estimates by b_S . Then the approximated estimates are given by

$$b_A = \frac{[\sum_{(m,n) \in \mathcal{C}} \pi^{(m,n)}] [\sum_{(m,n) \in \mathcal{C}} b_S^{(m,n)} \pi^{(m,n)}]}{1 - [\sum_{(m,n) \in \mathcal{C}} \pi^{(m,n)}]}, \quad (19)$$

and their weights are all equal as well, namely $1/(T-S)$. From Equations (18) and (19) we see that estimates of more important samples contribute more to the approximates. The second method thus uses not only closely related information (neighboring estimates), but also less related information (non-neighboring estimates). It is not a priori clear whether this is good or bad, and we shall investigate the issue below.

5.3 Comparison of the methods

We now have four methods for the approximation procedure, as follows:

Sampling method	Approximating method	
	Ave. neighbor	Norm. probability
Non-probability	Method 1	Method 2
Probability	Method 3	Method 4

We compare the four methods from two aspects: approximation accuracy and computation time. For approximation accuracy our criterium is the average absolute deviation from the true values. Since the non-probability methods (Methods 1 and 2) do not work if measurements in every group have equal prior probabilities, we randomly assign probabilities to variables in this case under the constraint that the probabilities in each group sum up to one. The performance of different methods is hardly affected by these random choices. (Further evidence is provided in our background document). We compare the four methods in four cases: (1) small number of samples, (2) moderate number of samples, (3) large number of samples, and (4) the full sample.

In each case, we obtain S by requiring $\alpha > \alpha^*$ and $S < S^*$ in Methods (1) and (2). Then, we keep the same S in Methods (3) and (4). Thus all methods are compared under the same number of samples. We choose S^* and α^* as follows:

	Case			
	(1)	(2)	(3)	(4)
S^*	3000	7000	40000	69120
α^*	0.30	0.50	0.85	1.00

The various methods and cases are compared in Table 6.

TABLE 6

Clearly, within each method, when we move from Case (1) to Case (4), the number of samples S and the precision α increase, and consequently the approximation accuracy is improved but computation time increases.

Comparing different approximating techniques, we find that Method 2 has higher approximation accuracy and needs less computation time than Method 1; and similarly that Method 4 has higher approximation accuracy and needs less computation time than Method 2. Apparently the normalization method strictly dominates the method using neighboring estimates, and this domination is especially strong when the number of samples is small.

Next, when we compare different sampling techniques, we see that no method strictly dominates the other. When the number of samples is small, Method 4 is more accurate than Method 2, but it is less accurate when the number of samples is large, thus reflecting the trade-off between using the more important estimates and a wider range of estimates. The computation time is higher for Method 4 than for Method 2, because randomness is time-consuming.

The convergence of the approximation accuracy for each of the four methods is given in Figure 2. Average absolute deviations decrease smoothly for non-probability methods, but less smoothly for probability methods because of the randomness. In Methods 1, 2, and 3, the average absolute deviations decrease sharply when the number of samples increases to around 5000, and when the number of samples exceeds 40000, adding more samples only leads to a marginal improvement of the accuracy. In Method 4 the decrease is slower. The accuracy of the approximated mean is similar to that of the approximated standard deviation in Methods 1 and 3, but much larger than the accuracy of the approximated standard deviation in Methods 2 and 4.

Figure 2: Approximation accuracy: 4 methods

This shows that normalization methods perform better than methods using neighboring estimates in the approximation of the standard deviation.

The computation time is roughly proportional to the number of samples, so that computation time can be accurately predicted for each method. In fact, the ratio

$$\frac{\text{Computation time (in seconds)}}{\text{number of samples}/100}$$

is approximately 1.5 (Method 3), 1.2 (Method 4), 1.0 (Method 1), and 0.7 (Method 2). In summary, Methods 2 and 4 dominate Methods 1 and 3. When the number of samples is relatively small, Method 4 is preferred, but when the number of samples is relatively large, then Method 2 is preferred.

6 Sensitivity with respect to the prior π

In Section 2.3 we distinguished between four levels of belief regarding the specification of the prior probabilities π . So far our empirical results are

based on the first level of belief (equal prior probability in each group). In this section we investigate the effects of the three other levels of belief of π on the estimates and standard deviations. In other words, we ask how sensitive the empirical results and the conclusions are with respect to π . We discuss the three levels in turn: unequal priors, ordered priors, and data-dependent priors.

6.1 Unequal priors

Suppose that in one group, say group l , the m_l variables do not all have the same prior probability $1/m_l$, but that one of the variables, say variable j , has a different probability π_l^j , while the remaining variables have equal probabilities

$$\pi_l^i = \frac{1 - \pi_l^j}{m_l - 1} \quad (i = 1, \dots, j - 1, j + 1, \dots, m_l). \quad (20)$$

This assumption can be made for any of the type I groups. For our sensitivity experiment we choose one focus variable ‘education’ and one auxiliary variable ‘democracy’. We choose education, because this group contains nine variables with relatively large deviations, and the effect of education on economic growth depends on which variable is used; see Table 4. Democracy is of interest because it has strong policy implications (Barro, 1999) and its effect on economic growth is controversial. Within education we choose the variable ‘primary schooling’ as the one whose prior probability is different, because its role in explaining economic growth appears to differ from other education variables; within democracy we choose the variable ‘political rights’.

In Figure 3 we report the sensitivity of four groups: education, democracy, health, and initial state. These groups are chosen because they are proximate determinants that are typically regarded as the most important growth theories. Also, by including education and democracy, we can investigate the direct effect (effect on the group itself) and the indirect effect (effect on other groups) of changing the prior probability. In the figure, the prior probability of primary schooling (top panel) and political rights (bottom panel) varies between 0 and 1. Note that the estimated means of all variables are linear with the varying prior probability, because π_l^i in Equation (20) is a linear function of π_l^j .

Changing the prior probability of primary schooling has a serious direct effect on the estimated mean and standard deviation of education. The estimated mean of education is negative when the prior probability of primary schooling is less than 0.2, but becomes positive when it is larger than 0.4.

Figure 3: Sensitivity with respect to π : unequal priors

This is due to the fact that primary schooling has a strong positive effect on growth while the effect of other education variables is weak.

The estimated standard deviation of education is a concave function of the prior probability. It is obvious that when primary schooling has weight 1, we obtain the minimal standard deviation because there is no variation between variables. It is less obvious that when primary schooling has a small weight, we also obtain small standard deviations. The reason lies in the fact that primary schooling differs much from other education variables, so that a small weight leads to a small cross term bb' in Equation (12).

The estimates in the other three groups (indirect effect) are less sensitive than those of education. The estimated means never change sign. Among the three groups, health appears to be the most sensitive to the change in prior. The estimated standard deviations of the three groups are all insensitive to

the change in prior.

In the bottom panel we change the prior on political rights in the group democracy. The effects are very small. Even the estimate of democracy itself (the direct effect) is not sensitive.

6.2 Ordered priors

Next suppose that we can order the priors in group l so that the priors of the variables $x_l^1, \dots, x_l^{m_l}$ are known to satisfy $\pi_l^1 > \dots > \pi_l^{m_l}$. In particular, assume that $\pi_l^{i+1} = r\pi_l^i$ for some $0 < r < 1$. Then,

$$\pi_l^i = \frac{(1-r)r^{i-1}}{1-r^{m_l}} \quad (i = 1, \dots, m_l). \quad (21)$$

The smaller is r the more weight is placed on the important variables. Equation (21) allows the prior probability of the most important measurement, π_l^1 , to change over the interval $(1/m_l, 1)$.

Figure 4 presents some representative examples when the priors are ordered. To perform this experiment we set $r = 1/2$ and we need a predetermined ordering of the variables. In the group education we select primary schooling as the most important variable and we order the other variables randomly. Unlike the previous case, neither the estimated mean nor the standard deviation is a linear function of the prior probability. Still, the main results are essentially the same as before. In particular, as the prior probability of primary schooling increases, the estimated mean of education changes from negative to positive, and the estimated standard deviation of education is a concave function. The health effect is weakened as the prior probability of primary schooling increases, while initial state and democracy are insensitive to the probability change.

In the group democracy we select political rights as the most important variable. In this group there are only two variables, so that $\pi^1 = 1/(1+r)$ and $\pi^2 = r/(1+r)$. Changing the probability hardly changes the estimated means and standard deviations for any of the groups.

We repeated these experiments for all other type I groups, both for the unequal prior case and for the ordered priors case. Based on these experiments we draw three conclusions regarding the sensitivity with respect to the prior probability. First, the effects of proximate determinants on economic growth is robust to the choice of prior probability, except for the group education. Second, the indirect effects of the prior probability are very small, while the direct effect varies across groups. The direct effect is large for those groups whose variables vary greatly, such as education. But it is small for

Figure 4: Sensitivity with respect to π : ordered priors

those groups whose variables are highly correlated, such as health, inflation, and scale effect. Third, the standard deviations are quite robust.

6.3 Data-dependent priors

In Section 2.3 we discussed two updating algorithms based on data-dependent priors: one-step updating and two-step updating.

TABLE 7

In Table 7 we present the updated priors and the new HWALS-F8 estimates for five groups: education, health, initial state, government intervention, and democracy. (Full results are in our background document.) By construc-

tion, the two methods give the same updated prior probabilities for the focus variables, but the methods differ in the computation of the updated prior probabilities for the auxiliary variables. However, the updated probabilities of the auxiliary variables and the resulting HWALS estimates are not much affected by which updating algorithm is used (except for demographic characteristics and scale effect).

The large variation in updated prior probabilities (ranging from 0.978 to 0.003 in the group education) shows that some variables in a group are much more relevant for economic growth than others. The ordering is generally in line with findings in other studies, e.g. SDM and Magnus et al. (2010). For example, the government consumption share is by far the most relevant variable in the government intervention group. Generally, the most relevant variables also have the highest posterior inclusion probability (SDM, 2004), or are the most significant (Magnus et al., 2010) compared to other variables in the same group.

Sometimes, variables in one group are almost equally relevant in explaining growth, and their updated prior probabilities are therefore close. Examples of such groups are: economy system, inflation, and war. The variables in these groups are highly correlated, and hence including all variables in one regression leads to very insignificant estimates for some or all of the variables. For example, both SDM (2004) and Magnus et al. (2010) report extremely weak correlation between capitalism/socialist dummy and economic growth. However, when considering them as a group, the correlation with growth is much stronger. Thinking in terms of groups rather than in terms of variables thus provides new insights.

The estimates produced by the two updating procedures generally have similar magnitudes and the same signs (except terms of trade ranking, Latin-American dummy, and fraction of land area near water). The exceptions all have a very weak effect on growth. The standard deviations are even more robust. Even though the two methods produce different updated prior probabilities for demographic characteristics and scale effect, the estimates and standard deviations in these two groups are close.

Finally, we compare the HWALS-F8 results after updating the priors (Table 7) with the equal probability default (Tables 4 and 5). There is a big difference between focus and auxiliary groups. In the focus groups (especially education), the effects are generally different and stronger when the priors are updated than in the equal probability case. In the auxiliary groups (such as democracy), the estimates and standard deviations when updating the priors are mostly in line with those using equal probabilities. The reason lies in the fact that all focus groups have a dominant variable. In summary, a different choice of priors π will lead to different HWALS estimates. The

difference between equal priors, unequal priors, and ordered priors is small. The effect of data-dependent priors is small for the auxiliary groups, but not small for the focus groups.

7 Grouping

It is not always easy to decide which variable belongs in which group. While the grouping in Tables 1 and 2 based on Durlauf et al. (2005) is plausible, there is no complete agreement in the growth literature on how to group the large number of growth proxies. Assigning a variable to a different group or separating one group into two will have an effect on the estimates. In Section 7.1 we investigate whether small changes in grouping have a small or a large effect on the estimates. Then, in Section 7.2, we discuss optimal grouping.

7.1 Sensitivity with respect to grouping

We consider four scenarios. First, we question whether the variable ‘public spending on education’ belongs in the group ‘education’. This variable has a low correlation with other education variables, so perhaps it should be placed in a separate group. This is scenario S1. Second, one might make a case for placing this variable in the group ‘government intervention’ (scenario S2). Third, we consider the possibility that enrollment rate and school years have a different effect on growth, because the former is a flow measure while the latter captures the quantity of education. Thus, in scenario S3, we separate enrollment rate (variables 5–7 and 8) from school years (variables 9–13). Finally, one may argue that the variables ‘years open’ and ‘outward orientation’ in the group ‘trade policy indices’ reflect different aspects of trade policy, and have a different effect on economic growth (Moon, 1998). In scenario S4 we place the two variables in separate groups.

TABLE 8

The four scenarios lead to eleven possible combinations, which we consider in Table 8. For each of the eleven combinations we report the average relative change between the HWALS-F8 estimates and standard deviations from the new grouping and the original grouping. The two columns labeled ‘all variables’ use all variables, the next two columns include only the focus variables, and the next two columns only the variables whose estimate (in absolute value) exceeds 0.001. The last column gives the number of sign changes over

all variables.

If we consider all variables, then some changes in grouping have a significant effect on the estimates, but the standard deviations are not much affected. This applies in particular to the case where public spending in education is separated from group ‘education’ (S1). Grouping public spending in education into government intervention (S2) affects the results only marginally. Separating enrollment rate and school years (S3) leads to moderate changes in the estimates and standard deviations. These changes show that different groupings of education makes a difference, but it does not tell us which option is preferred. We investigate this issue in the next subsection. Separating outward orientation and years open (S4) has only a weak effect and results in no sign changes. If we combine scenarios then the results become stronger, as expected. Upon closer inspection we see that the new grouping mainly leads to (direct) changes in the group itself, while (indirect) changes in other variables are much smaller. We also note that the new groupings lead to sign changes for no more than one variable in all cases. Both facts provide evidence of the robustness of our procedure.

Next we repeat these comparisons, but now only for the focus variables. Then the sensitivity is much reduced, showing that the auxiliary variables are less robust than the focus variables with respect to grouping. The large change in the auxiliary estimates is caused by the fact the effect of some of the auxiliary variables on growth is fragile, and hence that a small (absolute) change in the estimate leads to a large relative change. To support this statement, we exclude the variables with an estimate smaller than 0.001 (in absolute value). Indeed, the relative change based on these variables is about half of that based on all variables.

7.2 Optimal grouping

We can *statistically* investigate the ‘optimal’ grouping by estimating the number of groups, thus providing an explanation of the sensitivity results and suggesting the grouping from a purely statistical viewpoint. We focus on the grouping of education. In particular, we compare S1, S2, S3, S1&S3 (placing enrollment rate, school years, and public education spending in three separate groups), and S2&S3 (placing enrollment rate and school years in two separate groups, and public education spending in the group ‘government intervention’) against our benchmark. We employ three statistics in common use to determine the number of groups: Caliński and Harabasz’s (1974) *CH*-statistic, Hartigan’s (1975) *H*-statistic, and Krzanowski and Lai’s (1988) *KL*-statistic. Both *KL* and *H* show that the benchmark is the optimal grouping, while *CH* slightly favors S1. Optimal grouping of ‘trade policy indices’

cannot be studied in this way because the three statistics do not apply to groups with only two members. We conclude that the robust and important groups are consistent with our benchmark. Large relative changes are mostly due to groups with fragile coefficients.

8 Conclusions

Applied researchers frequently encounter the situation where there is more than one measurement (variable) for a concept (group). To include all variables of the group into the regression is not satisfactory, because of multicollinearity. To choose between variables based on diagnostics leads to pretesting. A satisfactory solution can be obtained through Bayesian model averaging, in particular two-level (hierarchical) Bayesian model averaging, where we question which groups should be in the model (level 1) and also which variables should be in each group (level 2). Our proposed method (HWALS) is an attempt to obtain estimates and standard deviations that fully reflect three sources of uncertainty: uncertainty represented by the error term, given the specification of the model; uncertainty about which (auxiliary) groups to include; and uncertainty about which variables to include in each group. Our method combines model selection and estimation and thus avoid the problem of pretesting. It is transparent, easy to implement, and computationally efficient compared to standard methods such as BMA and BACE. The method provides estimates and standard deviations for each group (concept) rather than several estimates corresponding to each variable (measurement), and this facilitates statistical inference and interpretation of the effect of the concept. Unlike factor analysis, HWALS allows clear economic explanations, because the data are not transformed (except for simple scaling).

We apply the HWALS theory to growth empirics, and study the effects of different growth theories in explaining cross-country growth. This application is particularly suitable, because the open-ended growth theories and the many possible proxies for the same concept expose growth regressions to a high degree of model uncertainty. The HWALS estimates appear to possess more intuitive signs and are generally more significant compared to other methods. For example, HWALS finds a moderately negative effect of democracy and fraction GDP in mining, which is theoretically justified and empirically supported by the literature, whereas BACE finds the opposite result. Our findings regarding the robust and important determinants are mostly in line with the literature. In particular, our results confirm that an East Asian dummy, and most proximate determinants, including initial

state, health, and price distortion (investment price) are the most robust and salient in explaining cross-country differences. Most of the weakly related determinants are also consistent with the growth theories and the previous literature. A notable difference from the literature, however, is that the education effect is not robust, which reflects the large variation between a wide range of education variables.

Extensive sensitivity analysis is provided with respect to the prior probabilities and grouping, from which we conclude that the main results, especially the estimates of robust and important determinants, are not sensitive. Also provided are methods of approximation when the number of groups or variables is large. We compare four approximation methods, and conclude that the approximation method using normalization has major advantages over the method using neighboring estimates, while non-probability and probability sampling reveal a trade-off between computation time and accuracy. The computation time of all methods increases linearly as the number of samples increases. The experimental results show that computation time can be much reduced while still obtaining estimates satisfying a given level of accuracy.

References

- R. J. Barro. Democracy and growth. *Journal of Economic Growth*, 1:1–27, 1996.
- R. J. Barro. Determinants of democracy. *The Journal of Political Economy*, 107:S158–S183, 1999.
- R. J. Barro and J.-W. Lee. International comparisons of educational attainment. *Journal of Monetary Economics*, 32:363–394, 1993.
- W. A. Brock and S. N. Durlauf. Growth empirics and reality. *World Bank Economic Review*, 15:229–272, 2001.
- T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3:1–27, 1974.
- D. Danilov and J. R. Magnus. On the harm that ignoring pretesting can cause. *Journal of Econometrics*, 122:27–46, 2004.
- G. De Luca and J. R. Magnus. Bayesian model averaging and weighted average least squares: equivariance, stability, and numerical issues. *The Stata Journal*, 11:518–544, 2011.

- J. P. Dunne and M. Uye. Military spending and development. Discussion Paper 0902, University of the West of England, Department of Economics, 2009.
- S. N. Durlauf, P. A. Johnson, and J. R. Temple. Growth econometrics. In S. N. Durlauf, editor, *Handbook of Economic Growth*, pages 555–677. North Holland, Amsterdam, 2005.
- S. N. Durlauf, A. Kourtellos, and C. M. Tan. Empirics of growth and development. In A. Dutt and J. Ros, editors, *International Handbook of Development Economics, Vol 1*, pages 32–47. Edward Elgar, Cheltenham, UK, 2008.
- C. Fernandez, E. Ley, and M. F. J. Steel. Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16:563–576, 2001.
- J. A. Hartigan. *Clustering Algorithms*. Wiley, New York, 1975.
- W. Krzanowski and Y. T. Lai. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44:23–34, 1988.
- E. Ley and M. F. Steel. Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics*, 29:476–493, 2007.
- J. R. Magnus and W. Wang. Concept-based Bayesian model averaging and growth empirics: Background document. Tilburg University, www.janmagnus.nl/items/magnus-wang-suppl.pdf, 2012.
- J. R. Magnus, O. Powell, and P. Prüfer. A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154:139–153, 2010.
- B. E. Moon. Exports, outward-oriented development, and economic growth. *Political Research Quarterly*, 51:7–36, 1998.
- X. Sala-i-Martin. I just ran two million regressions. *American Economic Review*, 87:178–83, 1997.
- X. Sala-i-Martin, G. Doppelhofer, and R. I. Miller. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 94:813–835, 2004.

Table 1: Grouping of variables: type I groups

<i>g</i>	Group	<i>v</i>	Variable
(1)	Demographic characteristics	1	Fraction population over 65
		2	Fraction population under 15
(2)	Economy system	3	Capitalism
		4	Socialist dummy
(3)	Education	5	Primary schooling (1960 enrollment rate)
		6*	Secondary schooling (1960 enrollment rate)
		7	Higher education (1960 enrollment rate)
		8	Public education spending share in GDP in 1960s
		9*	Primary school years
		10*	Secondary school years
		11*	Higher education years
		12*	Average years of schooling
		13*	Average years of schooling $\times$ log of GDP per capita
(4)	Government intervention	14	Public investment share
		15*	Public consumption share (excl. education and defense)
		16	Government consumption share in 1960s
		17	Government share of GDP in 1960s
		18	Nominal government GDP share in 1960s
(5)	Health	19	Life expectancy in 1960
		20	Malaria prevalence in 1960s
(6)	Inflation	21	Average inflation 1960–1990
		22	Square of inflation 1960–1990
(7)	Initial state	23	GDP per capita in 1960 (log)
		24	Size of economy (GDP in 1960)
(8)**	Democracy	25	Political rights
		26	Civil liberties
(9)	Scale effect	27	Land area
		28	Population in 1960
(10)	Trade policy indices	29	Outward orientation
		30	Years open
(11)	Tropics effect	31	Fraction of tropical area
		32	Tropical climate zone
		33	Absolute latitude
(12)	War	34	Fraction spent in war 1960–1990
		35	War participation 1960–1990

* Variable is not in the SDM (2004) data set, but taken from Sala-i-Martin (1997).

** Group (8) is called ‘Democracy’ following Barro (1999).

Table 2: Grouping of variables: type II groups

<i>g</i>	Group	<i>v</i>	Variable
(13)	Ethnicity and language	36*	Ethnolinguistic fractionalization
		37	English-speaking population
		38	Fraction speaking foreign language
(14)	Religion	39	Fraction Confucian
		40	Fraction Muslim
		41	Fraction Buddhist
		42	Fraction Protestant
		43	Fraction Hindu
		44	Fraction Catholic
		45	Fraction Orthodox
		46*	Religious intensity
(15)	Trade statistics	47*	Openness measure 1965–1974
		48	Primary exports in 1970
(16)	Terms of trade	49	Terms of trade ranking
		50	Terms of trade growth in 1960s
(17)	Regional effect	51	East Asian dummy
		52	African dummy
		53	European dummy
		54	Latin-American dummy
		55	Colony dummy
		56	British colony
		57	Spanish colony
(18)	Natural resource	58	Hydrocarbon deposits in 1993
		59	Fraction GDP in mining
		60	Oil-producing country dummy
(19)	Population	61	Population density coastal in 1960s
		62	Interior density
		63	Fraction population in tropics
		64	Population density in 1960
		65	Population growth rate 1960–1990
		66	Fertility in 1960s
(20)	Geography (excl. tropics effect)	67	Fraction land area near navigable water
		68	Landlocked country dummy
		69	Air distance to big cities
(21)	Price distortion	70	Investment price
(22)	Real exchange rate	71	Real exchange rate distortions
(23)	Defense	72	Defense spending share
(24)	Political instability	73	Revolutions and coups
(25)	Independence	74	Timing of independence

* Representative variable of the group.

Table 3: Within-group correlations

g	Group	v	Variable	Correlation
(1)	Demographic characteristics	1*	Fraction population over 65	
		2	Fraction population under 15	−0.91
(2)	Economy system	3	Capitalism	−0.58
		4*	Socialist dummy	
(5)	Health	19	Life expectancy in 1960	−0.73
		20*	Malaria prevalence in 1960s	
(8)	Democracy	25	Political rights	−0.83
		26*	Civil liberties	
(11)	Tropics effect	31	Fraction of tropical area	−0.89
		32	Tropical climate zone	−0.60
		33*	Absolute latitude	

* Adjusted variable.

Table 4: HWALS and WALS estimates: focus variables

Variable	WALS-F8	HWALS-F1	HWALS-F8
Education		−0.0013 (0.0038)	−0.0013 (0.0046)
5 Primary schooling	0.0037 (0.0188)		
6 Secondary schooling			
7 Higher education	−0.0079 (0.0081)		
8 Public edu. spending	−0.0007 (0.0160)		
9 Primary school yrs			
10 Secondary school yrs			
11 Higher education yrs			
12 Ave. school yrs			
13 Ave. school yrs $\times$ logGDP			
Health		0.0045 (0.0044)	0.0073 (0.0058)
19 Life expectancy	0.0144 (0.0271)		
20 Malaria prevalence	−0.0045 (0.0094)		
Initial state		−0.0030 (0.0046)	−0.0045 (0.0064)
23 GDP in 1960 (log)	−0.0073 (0.0168)		
24 Size of economy	0.0006 (0.0186)		
Tropics effect		−0.0029 (0.0032)	−0.0030 (0.0034)
31 Frac. of tropical area	0.0015 (0.0207)		
32 Tropical climate zone	0.0013 (0.0047)		
33 Absolute latitude	0.0054 (0.0195)		
Ethnicity and language			
36 Ethnolinguistic frac.	−0.0019 (0.0087)	−0.0024 (0.0025)	−0.0030 (0.0028)
37 English-speaking pop.	0.0014 (0.0053)		
38 Frac. foreign language	0.0006 (0.0062)		
Religion			
39 Fraction Confucian	0.0009 (0.0058)		
40 Fraction Muslim	−0.0004 (0.0079)		
41 Fraction Buddhist	0.0010 (0.0132)		
42 Fraction Protestant	−0.0122 (0.0161)		
43 Fraction Hindu	0.0003 (0.0074)		
44 Fraction Catholic	−0.0130 (0.0226)		
45 Fraction Orthodox	−0.0014 (0.0029)		
46 Religious intensity	−0.0035 (0.0095)	−0.0009 (0.0017)	−0.0015 (0.0019)
Price distortion			
70 Investment price	−0.0047 (0.0076)	−0.0029 (0.0015)	−0.0041 (0.0017)

Table 5: HWALS and WALS estimates: auxiliary variables

Variable	WALS-F8	HWALS-F1	HWALS-F8
Demographic characteristics		0.0026 (0.0047)	0.0027 (0.0048)
1 Frac. pop. over 65	−0.0011 (0.0204)		
2 Frac. pop. under 15	−0.0003 (0.0324)		
Economy system		−0.0010 (0.0016)	−0.0010 (0.0016)
3 Capitalism	0.0018 (0.0056)		
4 Socialist dummy	−0.0000 (0.0067)		
Government intervention		−0.0004 (0.0021)	−0.0003 (0.0021)
14 Public investment share	0.0016 (0.0044)		
15 Public consumption share (excl. education and defense)			
16 Gov. consumption share	−0.0367 (0.1602)		
17 Gov. share of GDP	0.0362 (0.1489)		
18 Nominal gov. GDP share	0.0001 (0.0078)		
Inflation		0.0005 (0.0022)	0.0005 (0.0022)
21 Average inflation	0.0042 (0.0179)		
22 Square of inflation	−0.0064 (0.0200)		
Democracy		0.0025 (0.0027)	0.0025 (0.0027)
25 Political rights	0.0047 (0.0102)		
26 Civil liberties	0.0002 (0.0075)		
Scale effect		0.0028 (0.0027)	0.0028 (0.0028)
27 Land area	0.0063 (0.0157)		
28 Population	0.0005 (0.0086)		
Trade policy indices		0.0010 (0.0025)	0.0009 (0.0024)
29 Outward orientation	−0.0008 (0.0055)		
30 Years open	−0.0032 (0.0104)		
War		−0.0001 (0.0017)	0.0001 (0.0016)
34 Frac. spent in war	0.0004 (0.0067)		
35 War participation	0.0022 (0.0086)		
Trade statistics			
47 Openness measure	−0.0004 (0.0147)	0.0007 (0.0029)	0.0006 (0.0029)
48 Primary exports	−0.0026 (0.0104)		
Terms of trade			
49 Terms of trade ranking	0.0028 (0.0084)	0.0004 (0.0027)	0.0004 (0.0027)
50 Terms of trade growth	0.0026 (0.0058)	0.0036 (0.0024)	0.0035 (0.0024)
Regional effect			
51 East Asian dummy	0.0087 (0.0108)	0.0062 (0.0028)	0.0058 (0.0028)
52 African dummy	0.0017 (0.0117)	−0.0033 (0.0035)	−0.0031 (0.0036)
53 European dummy	0.0198 (0.0247)	0.0016 (0.0045)	0.0015 (0.0045)
54 Latin-American dummy	0.0125 (0.0258)	−0.0012 (0.0045)	−0.0014 (0.0046)
55 Colony dummy	−0.0023 (0.0155)	−0.0040 (0.0035)	−0.0038 (0.0035)
56 British colony	−0.0003 (0.0071)	0.0028 (0.0027)	0.0028 (0.0027)
57 Spanish colony	−0.0015 (0.0152)	0.0015 (0.0033)	0.0012 (0.0033)
Natural resource			
58 Hydrocarbon deposits	0.0015 (0.0053)	0.0001 (0.0019)	0.0001 (0.0019)
59 Frac. GDP in mining	−0.0016 (0.0072)	−0.0014 (0.0019)	−0.0013 (0.0019)
60 Oil country dummy	−0.0020 (0.0052)	−0.0018 (0.0023)	−0.0018 (0.0023)

Table 5: Continued

Population			
61 Population density coastal	0.0019 (0.0172)	0.0010 (0.0029)	0.0007 (0.0030)
62 Interior density	-0.0025 (0.0070)	-0.0011 (0.0017)	-0.0010 (0.0017)
63 Fraction pop. in tropics	0.0003 (0.0092)	0.0014 (0.0032)	0.0015 (0.0032)
64 Population density	-0.0032 (0.0060)	-0.0016 (0.0021)	-0.0015 (0.0021)
65 Population growth rate	0.0073 (0.0232)	0.0013 (0.0053)	0.0014 (0.0054)
66 Fertility	0.0007 (0.0224)	-0.0031 (0.0061)	-0.0030 (0.0063)
Geography (excl. tropics effect)			
67 Frac. land area near water	0.0018 (0.0118)	0.0018 (0.0032)	0.0016 (0.0032)
68 Landlocked country dummy	0.0027 (0.0040)	0.0003 (0.0018)	0.0002 (0.0018)
69 Air distance to big cities	0.0009 (0.0102)	0.0010 (0.0025)	0.0010 (0.0025)
Real exchange rate			
71 Real exchange rate dist.	-0.0031 (0.0107)	-0.0024 (0.0020)	-0.0021 (0.0020)
Defense			
72 Defense spending share	-0.0145 (0.0599)	-0.0003 (0.0017)	-0.0004 (0.0017)
Political instability			
73 Revolutions and coups	0.0043 (0.0064)	-0.0005 (0.0019)	-0.0006 (0.0018)
Independence			
74 Timing of independence	-0.0001 (0.0110)	0.0006 (0.0025)	0.0008 (0.0025)

Table 6: Comparison between approximation methods

	S	α	Ave. of absolute deviations ($\times 10^{-4}$)		Comp. time (sec)
			Mean	Std	
Method 1					
Case (1)	2508	0.3029	8.8440	7.3064	55.81
Case (2)	6496	0.5050	4.9325	4.5899	94.03
Case (3)	26428	0.8663	0.6005	0.7716	318.89
Case (4)	69120	1.0000	0.0000	0.0000	680.70
Method 2					
Case (1)	2508	0.3029	1.9909	0.8846	21.13
Case (2)	6496	0.5050	1.3630	0.5483	48.72
Case (3)	26428	0.8663	0.4543	0.1539	204.84
Case (4)	69120	1.0000	0.0000	0.0000	485.67
Method 3					
Case (1)	2508	0.1445	3.2941	2.8367	66.23
Case (2)	6496	0.3263	2.3358	2.9017	122.28
Case (3)	26428	0.7855	0.9740	1.3896	425.31
Case (4)	69120	1.0000	0.0000	0.0000	1014.70
Method 4					
Case (1)	2508	0.1389	1.5354	0.6524	33.61
Case (2)	6496	0.3233	1.2089	0.5043	81.81
Case (3)	26428	0.7846	0.5230	0.1903	330.75
Case (4)	69120	1.0000	0.0000	0.0000	819.53

Table 7: HWALS estimates using data-dependent priors: selected groups

Variable	One-step updating		Two-step updating	
	HWALS-F8	updated π	HWALS-F8	updated π
Education	0.0051 (0.0034)		0.0050 (0.0034)	
5 Primary schooling		0.9784		0.9784
6 Secondary schooling		0.0033		0.0033
7 Higher education		0.0024		0.0024
8 Public edu. spending		0.0027		0.0027
9 Primary school yrs		0.0025		0.0025
10 Secondary school yrs		0.0031		0.0031
11 Higher education yrs		0.0019		0.0019
12 Ave. school yrs		0.0024		0.0024
13 Ave. school yrs $\times$ logGDP		0.0033		0.0033
Government intervention	0.0005 (0.0020)		0.0004 (0.0020)	
14 Public investment share		0.1555		0.1647
15 Public consumption share (excl. education and defense)		0.1710		0.0220
16 Gov. consumption share		0.4317		0.5210
17 Gov. share of GDP		0.0827		0.0440
18 Nominal gov. GDP share		0.1590		0.2484
Health	0.0062 (0.0059)		0.0065 (0.0060)	
19 Life expectancy		0.8142		0.8142
20 Malaria prevalence		0.1858		0.1858
Initial state	-0.0084 (0.0057)		-0.0082 (0.0059)	
23 GDP in 1960 (log)		0.6923		0.6923
24 Size of economy		0.3077		0.3077
Democracy	0.0015 (0.0024)		0.0017 (0.0024)	
25 Political rights		0.3017		0.2217
26 Civil liberties		0.6983		0.7783

Table 8: Sensitivity with respect to grouping

Grouping		Ave. relative change (%)						Number of variables with sign change
		All variables		Focus variables		Variables with $ b \geq 0.001$		
		Mean	Std	Mean	Std	Mean	Std	
(1)	S1	30.65	0.01	12.64	0.04	15.03	0.02	1
(2)	S2	6.35	0.60	4.78	0.78	3.19	0.57	0
(3)	S3	15.40	2.93	13.62	3.31	9.11	2.79	0
(4)	S4	12.19	3.06	1.52	1.95	4.59	2.68	0
(5)	S1 & S3	38.13	5.11	13.00	7.08	15.99	4.95	1
(6)	S1 & S4	35.99	5.86	12.93	3.94	15.59	5.50	1
(7)	S2 & S3	20.71	3.28	8.32	6.56	11.72	3.48	1
(8)	S2 & S4	14.71	3.18	5.42	2.25	5.56	2.92	0
(9)	S3 & S4	18.10	6.20	9.58	6.52	8.15	5.91	0
(10)	S1 & S3 & S4	43.74	8.59	14.93	9.70	17.61	8.60	1
(11)	S2 & S3 & S4	24.23	5.82	10.11	7.73	11.44	5.69	1

Explanation of grouping:

S1: ‘Public spending in education’ (in group ‘education’) placed in separate group

S2: ‘Public spending in education’ moved to group ‘government intervention’

S3: Group ‘education’ split in two groups: ‘enrollment rate’ (variables 5–8) and ‘school years’ (variables 9–13)

S4: The two variables ‘outward orientation’ and ‘years open’ in group ‘trade policy indices’ placed in separate groups