Concept-Based Bayesian Model Averaging and Growth Empirics*

JAN R. MAGNUS† and WENDUN WANG‡

† *Department of Econometrics and Operations Research, VU University Amsterdam, The Netherlands (e-mail: jan@janmagnus.nl)*

‡ *Econometric Institute, Erasmus University Rotterdam, The Netherlands (e-mail: wangwendun@gmail.com)*

*We are grateful to participants at seminars at Tilburg University, Erasmus University Rotterdam, Groningen University, and the 23rd $(EC)^2$ conference in Maastricht; to Gerda Claeskens, John Einmahl, and Arthur van Soest; and to the editor and two referees for constructive and helpful comments. JEL Classification numbers: C51, C52, C13, C11.

Abstract: In specifying a regression equation, we need to specify 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 determinants taking explicit account of the measurement problem in the growth regressions. We find that estimates produced by HWALS provide intuitive and robust explanations. We also consider approximation techniques which are useful when the number of variables is large or when computing time is limited.

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

I. Introduction

In applied econometrics, when estimating a regression equation, one has to decide which *concepts* (say inflation) to include in the regression: the 'specification' problem. In addition, one has to decide which *measurements* of these concepts to use (for example, CPI-based or PPI-based inflation): the 'measurement' problem. The measurement problem is common in practice because most economic variables can be measured in various ways. Climate, for example, as a potential determinant of growth, can be measured by the fraction of a country lying in the tropics, the area of a country lying in the tropics, or absolute latitude. Another example is the concept of market concentration, typically thought of as a factor that affects the financial stability of individual firms, which can be measured by the Herfindahl-Hirschman index but also by the market share of, say, the four largest firms.

The measurement problem raises at least three issues. First, different choices of measurements produce different estimates for the same concept, leading to ambiguity in explanation and policy implications. Second, multiple measurements typically cause multicollinearity if they are included in one regression model, so that the estimates for individual measurements lack precision, and statistical inference on a concept based on these estimates can therefore be misleading. Third, including multiple measurements in one model can also cause a problem of dimensionality when the number of explanatory variables is close to or even exceeds the number of observations.

The current paper addresses the measurement problem 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 a method called hierarchical weighted least squares

(HWALS), a generalization of weighted average least squares (WALS) developed in 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 problems associated with pretesting (see Danilov and Magnus (2004) for a discussion and review of these problems); it allows researchers to assign various types of priors depending on the strength of their information and beliefs; it limits the extent of multicollinearity and dimensionality problems because it only considers models with one variable in each group; and its computational burden is very light, especially compared to standard Bayesian model averaging (BMA) and Bayesian averaging of classical estimates (BACE).

In the empirical growth literature the three types of uncertainty are especially important, because there is little consensus in this literature 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. 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) listed 145 candidate variables, while the number of countries is typically less in cross-country growth studies. Our paper employs HWALS to re-investigate the effects of various growth determinants.

We mainly compare our estimates with those of Sala-i-Martin *et al.* (2004) and

with the WALS estimates of Magnus *et al.* (2010). 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. Our empirical results also provide several new insights. For example, we find — in contrast to the current literature — that education and relative government size (government's economic activities) are not robust, because some of the variables in these groups have poor explanatory power in the growth regressions.

The paper is organized as follows. A literature review is provided in Section II. In Section III we present the hierarchical estimation strategy. Section IV describes the data, grouping, and scaling. We apply our estimation strategy to the data in Section V 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 VI. Section VII concludes. In our supplementary document (Magnus and Wang, 2014) we present extensions and more detailed analyses.

II. A brief review of the literature

The measurement problem is not new — it was mentioned, *inter alia*, in Brock *et al.* (2003) in the context of growth empirics. A popular method to deal with it is 'extreme bounds analysis' (Leamer and Leonard, 1983; Leamer, 1985), but this method has the disadvantage (in contrast to HWALS) that it produces various estimates for each concept. Another conventional method, called 'pretesting', is to try many different concepts and select the most appealing combination. There are many problems with this procedure (Danilov and Magnus, 2004) caused by the fact that model selection

and estimation are completely separated, so that uncertainty in the model selection is ignored when reporting properties of the estimates. In contrast to pretesting, HWALS combines model selection and estimation in one procedure.

One may also employ a factor-augmented regression model. Here we must decide on the number of factors (pretest problem), and when more than one factor is used the explanation of a concept becomes more difficult. A possible solution to both problems is to extract just one factor from each group, but then we would use only a small portion of the information in the data.

Since Raftery *et al.* (1997), Bayesian model averaging has developed as a popular tool in addressing model uncertainty, especially in the application to the empirical growth literature; see, *inter alia*, Fernández *et al.* (2001), Sala-i-Martin *et al.* (2004), and Ciccone and Jarociński (2010). Standard Bayesian model averaging addresses model uncertainty (which concepts to include) in growth regressions, while our approach addresses both model uncertainty and measurement uncertainty simultaneously, in the same spirit as Salimans (2012) who studied functional-form uncertainty and model uncertainty simultaneously.

A recent study by Durlauf *et al.* (2008b) investigated the robustness of growth theories using Bayesian model averaging with a dilution prior. This is related to what we do, although the growth theories and their empirical proxies studied in Durlauf *et al.* (2008b) differ in an essential way from our 'concepts' and 'measurements'. Multiple empirical proxies capture different aspects of a growth theory, and each aspect itself is a concept. For example, Durlauf *et al.* (2008b) considered two proxies for the geography theory, namely the fraction of tropical/subtropical land area and the fraction of land near navigable water. These two proxies indeed measure two effects of geography on growth: climate and physical accessibility (two concepts),

and for each concept they only have one measurement. The two proxies of the geography theory are not alternative measurements for the same concept (the correlation is only around 0.14), and this is where their paper differs from ours. In our case, standard BMA cannot be applied, primarily because we do not allow our model space to contain models with multiple measurements within one group. The use of a dilution prior (George, 2010) captures the dilution property resulting from multicollinear variables, but it does not address the fact that multiple measurements of a concept are included in one model, leading to misleading Bayesian model averaging estimates (due to misleading likelihoods and estimates obtained from models containing multicollinear variables in the same group). By shrinking the model space, our HWALS procedure addresses this problem and also reduces the computational burden. HWALS thus also differs from the hierarchical dilution prior used in Durlauf *et al.* (2012) who worked with the whole model space.

Our approach is also related to the jointness statistic proposed in Doppelhofer and Weeks (2009), which measures the dependence between explanatory variables. The jointness measure is the posterior probability that two or more variables appear in the same model. Multiple measurements of a concept are correlated with each other and are likely, but not certain, to have strong negative jointness. Conversely, variables that have negative jointness do not necessarily measure the same concept. Like other Bayesian approaches, the jointness measure computed from the posterior probability is also affected by the multicollinearity of variables in the same group.

Our work is in the same spirit as the hierarchical structure studied by Brock *et al.* (2003) and the heredity prior proposed by Chipman (1996). Brock *et al.* (2003) employed a tree structure to construct prior probabilities, while Chipman (1996) considered priors for group predictors and for competing predictors. Our hierarchical

averaging method resembles these two approaches, especially since all three methods average over a subset of models. But our method differs from the two approaches in at least four aspects. First, unlike Brock *et al.* (2003) who assigned equal and independent weights to each growth theory in a tree structure, HWALS allows for inequality and dependence between the various theories. Second, compared with the heredity prior, the method of restricting the model space is much simpler in HWALS (groups and variables). Third, our procedure allows us to assign various types of priors to measurements (imprecise priors, data-dependent priors) depending on the strength of the researcher's beliefs. Finally, HWALS provides an explicit form of the first two unconditional moments.

III. Hierarchical weighted average least squares

Groups and variables

We write the linear regression model as

$$y = X^*\beta^* + \epsilon = X_1^*\beta_1^* + X_2^*\beta_2^* + \epsilon, \tag{1}$$

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

$$X_1^* = \left(x_{1,1}^*, \ldots, x_{1,k_1}^*\right), \qquad X_2^* = \left(x_{2,1}^*, \ldots, x_{2,k_2}^*\right), \tag{2}$$

and the components of the (group) parameter vectors as

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

The distinction between *groups* and *variables* is important. 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. 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,

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

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

$$\sum_{i=1}^{m_{1,l_1}} \pi_{1,l_1}^i = 1, \qquad \sum_{j=1}^{m_{2,l_2}} \pi_{2,l_2}^j = 1. \tag{5}$$

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

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

and the parameter vectors

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

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

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

where we emphasize again that each model includes precisely *one* variable from each group.

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) and (j) is given by

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

The validity of the independence assumption embodied in (9) depends on how the groups are set up, and it is therefore important to investigate the sensitivity of the results to different groupings. We consider this issue in Section V. 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. A major

advantage of Bayesian model averaging is that it treats model selection and estimation as *one* procedure, thus incorporating not only the error uncertainty but also the model uncertainty. We shall use a method called WALS (weighted average least squares), but this is not essential in the development. WALS is a model averaging approach, taking an intermediate position between Bayesian and frequentist methods. It averages the estimates obtained from constrained least squares (frequentist) but introduces Bayesian components in the weighting scheme. The advantages of WALS are both conceptual and computational. The choice of priors in WALS mimics ignorance (the typical situation in model specification) and is also near-optimal in the sense of minimizing some risk or regret criterion. The computation time of WALS is negligible (of order k_2) due to a semi-orthogonal transformation of the auxiliary variables, in contrast to the heavy computational burden of standard BMA and frequentist model averaging (of order 2^{k_2}).

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). The prior chosen is Laplace, although robust versions now exist (Kumar and Magnus, 2013). The WALS procedure was recently reviewed in Magnus and De Luca (2014), where the reader can find elaborate discussions of the advantages and disadvantages of this model averaging procedure.

We thus obtain the posterior mean (the WALS estimates),

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

and the posterior variance matrix $V^{(i,j)}$. This is the second step.

These posterior moments are, of course, still conditional on the choice of variables, that is, on (i) and (j). In the third and final step we obtain the unconditional

posterior moments b and V from

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

and

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

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 more different variables in a group, the larger V). The estimator b is the hierarchical WALS (HWALS) estimator, and V is taken to be its variance.

The HWALS estimator b cannot be interpreted as the usual marginal effect, since it corresponds to a group (concept) rather than to a variable (measurement). Since all variables are normalized to the same scale (see Section IV), the estimated coefficient of the i-th variable in a group is the *normalized* marginal effect, taking into account possible inclusion of other auxiliary variables. Due to the normalization, such effects are comparable not only within concepts but also between concepts. The averaged estimator (over the variables) of a group coefficient can thus be interpreted as the average effect of the group. Like other model averaging estimators the HWALS estimator belongs to the class of biased estimators, and hence t-ratios do not have their usual interpretations. One limitation of HWALS is that standard statistical tests of coefficients are not possible in HWALS since the distribution of WALS estimates is not known.

Choice of π

The prior probabilities π should be specified, and the question is how. The specification of π should depend on the strength of the researcher's prior information and beliefs on the 'quality' of the variables. We distinguish between four cases.

In the first case, the researcher has no prior information at all. In each group the quality of one variable is 'independent' of the quality of another, and equally good, so we assign equal weights within each group, that is,

$$\pi^i_{1,l_1} = \frac{1}{m_{1,l_1}}, \qquad \pi^i_{2,l_2} = \frac{1}{m_{2,l_2}}.$$

This is our default. Discrete uniform priors (over models) in a Bayesian model averaging framework were recently criticized by George (2010), especially in the presence of highly correlated regressors. He suggested the use of dilution priors in order to prevent the probability of a set of 'similar' models increasing when more similar variables are introduced. While this is a good idea, our case is different, because our prior probabilities are assigned to variables rather than the models, and thus the probabilities are not diluted by highly correlated variables.

In the second case, the researcher has no prior information but hopes to update the prior using the observed data. 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

$$y = X_{11}\beta_{11} + \beta^i_{1,l}x^i_{1,l} + \epsilon \qquad (i = 1, \ldots, m_{1,l}), \tag{13}$$

from which we calculate the likelihood $L(x^i_{1,l}) = \Pr\left(y, X | x^*_{1,l} = x^i_{1,l}\right)$. Then we update the prior $\pi^i_{1,l}$ by Bayes' rule:

$$\bar{\pi}^i_{1,l} = \Pr\left(x^*_{1,l} = x^i_{1,l} | y, X\right) = \frac{\pi^i_{1,l} L(x^i_{1,l})}{\sum_{h=1}^{m_{1,l}} \pi^h_{1,l} L(x^h_{1,l})}.$$

A larger weight is thus assigned to the variable with more explanatory power (larger likelihood). Equation (13) is misspecified, because we ignore X_{12}^* (except one variable $x_{1,l}^i$) and all auxiliary regressors in X_2^*. However, the effect of the misspecification on $\bar{\pi}$ is partially 'divided out' and thus expected to be small. We confirmed this expectation by randomly including some additional controls in the regressions; see Magnus and Wang (2014) for details.

Two subcases are of interest. In case 2(a) (one-step updating) we update the priors for the auxiliary variables in the same way, based on the equation

$$y = X_{11}\beta_{11} + \beta_{2,l}^j x_{2,l}^j + \epsilon \qquad (j = 1, \ldots, m_{2,l}). \tag{14}$$

In case 2(b) (two-step updating) we update the priors for the auxiliary variables based on the extended equation

$$y = X_1^{(i)}\beta_1^{(i)} + \beta_{2,l}^j x_{2,l}^j + \epsilon \qquad (j = 1, \ldots, m_{2,l}), \tag{15}$$

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 $\bar{\pi}_{1,l}^i$.

The third case occurs when we have unequal prior information about the variables, and the exact values of prior probabilities are also known.

In the fourth case we can rank the prior probabilities within one group without knowing their precise values. Here we use 'imprecise probability' as our prior, namely $[\pi^-, \pi^+]$. This generalization of precise (point-valued) probability satisfies all principles of probability theory (Walley and Fine, 1982; Weichselberger, 2000), and allows us to model the uncertainty of subjective prior probabilities. The resulting estimates b and V are then also interval-valued.

IV. Data, grouping, and scaling

We reexamine growth determinants using the proposed hierarchical method of Section III. There is a large literature on explaining cross-country growth differences, but this literature has not led to a consensus on which determinants to include and which measure of each determinant to use. These issues are well exposed in Brock *et al.* (2003). 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 measure of each determinant to use.

Our data are taken primarily from Sala-i-Martin *et al.* (2004). The dependent variable is the average growth rate of GDP per capita 1960–1996. The Sala-i-Martin *et al.* (2004) 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 relative government size. 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 cannot estimate the whole set. Grouping will therefore be especially helpful here. The issue of having more variables than observations has recently received new attention in the literature; see Huang *et al.* (2010) and Jensen and Würtz (2012) for alternative approaches.

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 we also split the 'government' group in two ('relative government size' and 'defense'). The reason is that within the new groups 'tropics effect' and 'relative government size' 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, the extent of democracy in a country (the concept) can be measured in several ways, and we allow two measurements (political rights and civil liberties). An 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, or the share of public education spending. These input variables serve as different (but typically highly correlated) measurements for the same concept. We want to use only *one* measurement, but we do not know which one. Our theory of Section III applies to this type, that is, to groups (1)–(12) in Table 1.

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. 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 select one representative; for groups (16)–(20) 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. 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. For example, one may argue that the enrollment rates and attainment levels in the education group may have different effects on growth, because the former relate to the flow of education (Mankiw *et al.*, 1992) whereas the latter refer to stocks. We address such problems in Section V.

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. The latter effect can be removed by scaling the matrix

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

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.

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. For example, in the 'health' group 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. The five variables that have been re-signed are the fraction of the population over 65, the socialism dummy, malaria prevalence, civil liberties, and absolute latitude. The within-group correlations are presented in Magnus and Wang (2014, Table 1).

V. 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 (Brock and Durlauf, 2001), 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).

A few words are in order to explain why some groups are chosen to be focus and some to be auxiliary. We 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. More information is used in the second variant, HWALS-F8, where eight groups (including the constant term) are treated as focus groups. These eight groups are thus included in every model. 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). 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). Since the estimates in these two variants are highly similar, we only report results for HWALS-F8.

The choice of focus variables is motivated by the robust determinants of the endogenous growth model, and are in line with the choices in Magnus *et al.* (2010). More particularly, education and health capture different facets of human capital, with the former as a direct measure and the latter as a proxy for non-educational human capital. The initial state controls the convergence of growth (see also Section V). These variables are thought to be the most established drivers in both the neoclassical and the endogenous growth models. Instead of using the equipment investment as a measure of physical capital accumulation, we follow Sala-i-Martin *et al.* (2004) and Magnus *et al.* (2010) and use the investment price (labeled as price distortion) as a variable for domestic investment. Our focus variables include three slowly-moving 'fundamental' growth determinants (see Durlauf *et al.* (2008a) and the references therein). Ethnicity and language together with religion represent the degree of fractionalization and the culture in a society. The tropics effect is one of the most important features of a country's geography because countries in the tropical zone possess production technology which is less modern than the technology used in more temperate zones (Sachs, 2000). These three determinants are shown to be the most salient and robust among other fundamental variables in a number of growth studies (Sala-i-Martin, 1997; Fernández *et al.*, 2001; Sala-i-Martin *et al.*, 2004; Magnus *et al.*, 2010). The eight focus variables are the same as in Magnus *et al.* (2010), allowing comparisons to be made.

TABLES 3 and 4

In Tables 3 and 4 we present the results for HWALS-F8 using uniform priors and data-dependent priors, and compare them with WALS-F8. The sensitivity of the results to using other priors is studied in Magnus and Wang (2014), where we also present the results for HWALS-F1. We find that the effects of proximate determinants on economic growth are robust to the choice of prior probability, except for the education group. The indirect effect (effect on other groups) of a different choice of prior probability is small, but the direct effect (effect on the group itself) varies across groups. In general, the choice of priors is not a serious issue for the estimation of the standard deviations in our growth empirics.

The WALS-F8 estimates are based on the 67 variables in Sala-i-Martin *et al.* (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 HWALS-F8 estimates to groups.

We shall also compare our results with the BACE estimates of Sala-i-Martin *et al.* (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 Sala-i-Martin *et al.* (2004). The full set of unconditional BACE estimates is available in Magnus and Wang (2014).

Sign comparisons

Let us first compare the signs of HWALS-F8 using uniform priors with those of WALS-F8 and BACE, where we recall that the last two methods are based on variables while HWALS is 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 five of the type I groups the estimates are partially different, and in three type I and nine type II 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 suggests that regions with higher fractions of tropical land have lower growth rates, while all variables capturing the effect of the tropics 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. The HWALS estimates reflect the fact that 45% of Latin American countries and 86% of Sub-Saharan African countries achieve growth rates below or around the first quartile, while 52% of British colonial countries achieve above-average growth rates.

For the education group, 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 education group. Therefore, the education effect on growth appears to be inconclusive, in line with current literature.

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. The signs of the HWALS estimates are mostly in line with the estimates produced by only including one variable for each concept in a regression (not multiple variables for one concept). For example, BACE produces opposite effects of political rights and civil liberties, while HWALS finds a negative effect ($b = 0.0025$, $t = 0.9$) of democracy (recall that civil liberties is re-signed, so that a positive sign implies a negative effect), supporting the argument of Barro (1996). The HWALS estimate is in line with the HWALS estimates of political rights ($b = 0.0028$, $t = 1$) and civil liberties ($b = 0.0022$, $t = 0.9$) if we include each variable separately. However, if both variables are included simultaneously, then we obtain much smaller estimates of both variables (0.0022 for political rights, 0.0011 for civil liberties) and large variances due to high correlation ($r = 0.8237$). Also in contrast to BACE, HWALS finds a positive correlation between growth and the European dummy, and concludes that a larger fraction of GDP in mining leads to a

lower growth rate, which is supported by most cross-country studies on the 'resource curse'. Finally, 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.

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 for this group of approximately 1. The estimate of 'ethnolinguistic fractionalization' 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 and type II growth determinants produced by HWALS are more precisely estimated than in WALS and BACE (for example, demographic characteristics, inflation, and the scale effect).

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 the relative price of investment ($b = -0.0041$, $|t| = 2.4$) and the East Asian dummy ($b = 0.0058$, $|t| = 2.1$) are the most robust variables and explain much of the cross-country variation. Less robust but stronger in explanatory power is health ($b = 0.0073$, $|t| = 1.3$). Even less robust but still strong in explanatory power are 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 (Fernández *et al.*, 2001; Sala-i-Martin *et al.*, 2004; Durlauf *et al.*, 2008b).

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 Sala-i-Martin *et al.* (2004) based on posterior inclusion probabilities: economic growth is not found to be robustly related to education or relative government size.

Data-dependent priors

The last column in Tables 3 and 4 presents HWALS-F8 estimates using data-dependent priors with one-step updating. (The updated priors and results of two-step updating are presented in Magnus and Wang (2014).) By construction, the two updating methods give the same updated prior probabilities for the focus variables, but they differ in the computation of the updated prior probabilities for the auxiliary variables. The differences are small except for demographic characteristics and the scale

effect. 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 robustness of the updated probabilities and the resulting estimates confirms that model specification only has a marginal effect in the updating procedure.

We compare the HWALS-F8 results after updating the priors with the equal probability default. 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. The reason lies in the fact that all focus groups have a dominant variable, while most auxiliary groups have equally important variables. For example, the large variation in updated prior probabilities (ranging from 0.978 to 0.003) in the education group shows that some variables in this group are much more relevant for economic growth than others. The ordering is generally in line with findings in other studies, e.g. Sala-i-Martin *et al.* (2004) and Magnus *et al.* (2010). Generally, the most relevant variables also have the highest posterior inclusion probability (Sala-i-Martin *et al.*, 2004), or are the most significant (Magnus *et al.*, 2010) compared to other variables in the same group.

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. As discussed above, this is because the variables in most auxiliary groups are almost equally important, so that their updated prior probabilities are close. The variables in these groups are highly correlated, and hence including all variables in one regression leads to very unprecise estimates for some or all of the variables. Thinking in

terms of groups rather than in terms of variables thus provides new insights.

Effect of different groupings

Our empirical results are based on the grouping in Tables 1 and 2. These groupings can of course be questioned and we briefly discuss the effect of alternative groupings. More detailed results are presented in our supplementary document (Magnus and Wang, 2014).

Initial state. In the 'initial state' group we separate the two variables GDP per capita in 1960 and the initial size of the economy, motivated by the neoclassical growth model where initial GDP per capita has a structural role and thus should always be included (Mankiw *et al.*, 1992). We thus treat GDP per capita in 1960 as a focus variable and the initial size of the economy as auxiliary. Since the initial level of income is now always included, the estimated coefficients should be interpreted as the effects of determinants of the height of the steady-state growth path, rather than as the effects of long-run growth determinants. The new grouping leads to an estimated coefficient of the initial level of income ($b = -0.0098$), which is much larger in absolute value and has a smaller variance ($V = 0.0053$), making initial income an important determinant and providing strong evidence of convergence. Results of other focus groups and most auxiliary groups are not greatly affected.

Education. Education is a difficult concept to measure and our grouping can be easily criticized. We discuss four alternative groupings:

(i) Separate public education spending from the education group;

(ii) Assign public education spending to the relative government size group;

(iii) Distinguish between education flows and stocks by separating enrollment rates, attainment levels, and public education spending in three groups; and

(iv) Distinguish between lower and higher education level by separating primary and secondary education, higher education, and public education spending in three groups.

The results confirm the large variation of education variables as well as their distinct effects on growth. Growth is only weakly related to various aspects of education (flows versus stocks, lower versus higher level), with the exception of primary schooling.

Tropics effect. Separating latitude from tropic effect group hardly affects the results.

VI. Approximations for large k

To compute the HWALS estimates we need many runs of the WALS algorithm. Each run 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 3 and 4 are based on exact computations.

If the number of groups and variables increases further, then estimating all combinations (i, j) 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 regressions (Equation (8)) from all combinations and obtaining the corresponding WALS estimates for each regression; and assigning estimates to the non-sampled regressions based on the estimates of the sampled regressions. We shall discuss each aspect in turn.

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

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

representing the sum of the prior probabilities of the exact estimates used in the approximated HWALS computation. 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 prior probabilities as weights and draws randomly (without replacement) based on these weights. Each combination 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 $S < S^*$.

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. Neighboring estimates are good approximations 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).

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

$$\pi_*^{(i,j)} = \frac{\pi^{(i,j)}}{\sum_{(m,n)\in\mathcal{C}} \pi^{(m,n)}}, \qquad (i,j) \in \mathcal{C}, \tag{16}$$

where $\mathcal{C}$ is the set of sampled combinations. From Equation (16) we see that estimates of more important samples contribute more to the approximations. 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.

Comparison of the methods

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

	Approximating method	
Sampling 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 criterion is the average absolute deviation from the estimates obtained from the whole sample.

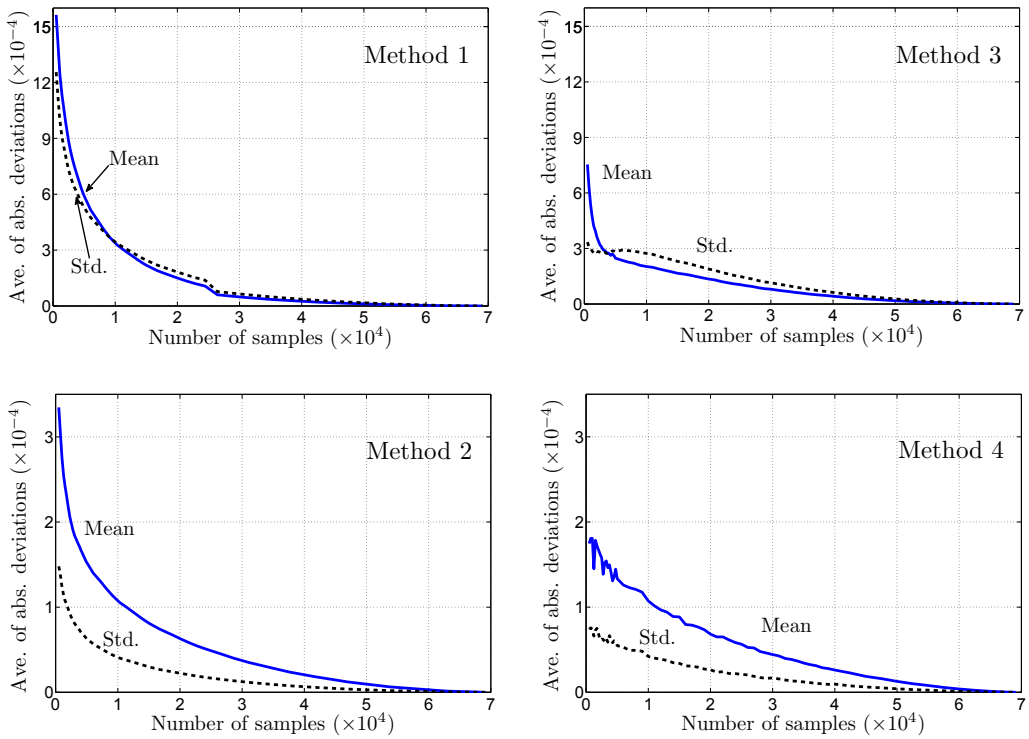

Figure 1: Approximation accuracy: four methods

Figure presents the convergence of the approximation accuracy for each of the four methods. Average absolute deviations decrease smoothly for non-probability methods, but less smoothly for probability methods because of the randomness. 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 3. 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 another. 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 roughly proportional to the number of samples, so that computation time can be 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). The computation time is higher for probability sampling than for non-probability sampling, because randomness is time-consuming.

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.

VII. 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 in the regression is not satisfactory, because of multicollinearity. To choose between variables based on diagnostics leads to pretesting problems. A satisfactory solution can be obtained through 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 avoids 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 one estimate and standard deviation 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. The hierarchical structure also allows us to assign various types of priors, depending on the strength of the researchers' beliefs. 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 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. Our findings regarding the robust and important determinants are mostly in line with the literature. A notable difference from the literature is that the education and relative government size effects are not robust, reflecting the large variation between variables in these two groups.

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. The experimen-

tal results show that computation time can be much reduced while still obtaining estimates satisfying a given level of accuracy.

Generalizations in various directions are possible. For example, non-linear models can be incorporated by adjusting the estimation method in the first-level averaging. The idea of hierarchical averaging can also be applied to other situations involving more than one level of uncertainty, such as model uncertainty with occasional structural breaks.

References

Barro, R. J. (1996). 'Democracy and growth', *Journal of Economic Growth*, Vol. 1, pp. 1–27.

Barro, R. J. (1999). 'Determinants of democracy', *The Journal of Political Economy*, Vol. 107, pp. S158–S183.

Barro, R. J. and Lee, J.-W. (1993). 'International comparisons of educational attainment', *Journal of Monetary Economics*, Vol. 32, pp. 363–394.

Brock, W. A. and Durlauf, S. N. (2001). 'Growth empirics and reality', *World Bank Economic Review*, Vol. 15, pp. 229–272.

Brock, W. A., Durlauf, S. N., and West, K. D. (2003). 'Policy evaluation in uncertain economic environments', *Brookings Papers on Economic Activity*, pp. 235–301.

Chipman, H. (1996). 'Bayesian variable selection with related predictors', *The Canadian Journal of Statistics*, Vol. 24, pp. 17–36.

Ciccone, A. and Jarociński, M. (2010). 'Determinants of economic growth: Will data tell?', *American Economic Journal: Macroeconomics*, Vol. 2, pp. 222–246.

Danilov, D. and Magnus, J. R. (2004). 'On the harm that ignoring pretesting can cause', *Journal of Econometrics*, Vol. 122, pp. 27–46.

De Luca, G. and Magnus, J. R. (2011). 'Bayesian model averaging and weighted average least squares: Equivariance, stability, and numerical issues', *The Stata Journal*, Vol. 11, pp. 518–544.

Doppelhofer, G. and Weeks, M. (2009). 'Jointness of growth determinants', *Journal of Applied Econometrics*, Vol. 24, pp. 209–244.

Durlauf, S. N., Johnson, P. A., and Temple, J. R. W. (2005). 'Growth econometrics', in S. N. Durlauf (ed.), *Handbook of Economic Growth*, North Holland, Amsterdam, pp. 555–677.

Durlauf, S. N., Kourtellos, A., and Tan, C. M. (2008a). 'Empirics of growth and development', in Dutt, A. K. and Ros, J. (eds), *International Handbook of Development Economics*, Vol. 1, Edward Elgar, Cheltenham, UK.

Durlauf, S. N., Kourtellos, A., and Tan, C. M. (2008b). 'Are any growth theories robust?', *The Economic Journal*, Vol. 118, pp. 329–346.

Durlauf, S. N., Kourtellos, A., and Tan, C. M. (2012). 'Is God in the details? A reexamination of the role of religion in economic growth', *Journal of Applied Econometrics*, Vol. 27, pp. 1059–1075.

Fernández, C., Ley, E., and Steel, M. F. J. (2001). 'Model uncertainty in cross-country growth regressions', *Journal of Applied Econometrics*, Vol. 16, pp. 563–576.

George, E. I. (2010). 'Dilution priors: Compensating for model space redundancy', in J. O. Berger, T. T. Cai, and I. M. Johnstone (eds), *Borrowing Strength: Theory Powering Applications — A Festschrift for Lawrence D. Brown*, Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 158–165.

Huang, J., Horowitz, J. L., and Wei, F. (2010). 'Variable selection in nonparametric additive models', *The Annals of Statistics*, Vol. 38, pp. 2282–2313.

Jensen, P. S. and Würtz, A. H. (2012). 'Estimating the effect of a variable in a high-dimensional linear model', *The Econometrics Journal*, Vol. 15, pp. 325–357.

Kumar, K. and Magnus, J. R. (2013). 'A characterization of Bayesian robustness for a normal location parameter', *Sankhya (Series B)*, Vol. 75, pp. 216–237.

Leamer, E. E. (1985). 'Sensitivity analyses would help', *American Economic Review*, Vol. 75, pp. 308–313.

Leamer, E. E. and Leonard, H. B. (1983). 'Reporting the fragility of regression estimates', *The Review of Economics and Statistics*, Vol. 65, pp. 306–317.

Magnus, J. R. and De Luca, G. (2014). 'Weighted-average least squares: A review', submitted for publication.

Magnus, J. R., Powell, O., and Prüfer, P. (2010). 'A comparison of two model averaging techniques with an application to growth empirics', *Journal of Econometrics*, Vol. 154, pp. 139–153.

Magnus, J. R. and Wang, W. (2014). 'Concept-based Bayesian model averaging and growth empirics: Supplementary document', available online at: www.janmagnus.nl/items/magnus-wang-suppl.pdf.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). 'A contribution to the empirics of economic growth', *The Quarterly Journal of Economics*, Vol. 107, pp. 407–437.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). 'Bayesian model averaging for linear regression models', *Journal of the American Statistical Association*, Vol. 92, pp. 179–191.

Sachs, J. D. (2000). 'Tropical underdevelopment', CID Working Papers No. 57, Center for International Development, Harvard University.

Sala-i-Martin, X. (1997). 'I just ran two million regressions', *American Economic Review*, Vol. 87, pp. 178–183.

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). 'Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach', *American Economic Review*, Vol. 94, pp. 813–835.

Salimans, T. (2012). 'Variable selection and functional form uncertainty in cross-country growth regressions', *Journal of Econometrics*, Vol. 171, pp. 267–280.

Walley, P. and Fine, T. L. (1982). 'Towards a frequentist theory of upper and lower probability', *Annals of Statistics*, Vol. 10, pp. 741–761.

Weichselberger, K. (2000). 'The theory of interval-probability as a unifying concept for uncertainty', *International Journal of Approximate Reasoning*, Vol. 24, pp. 149–170.

TABLE 1

Grouping of variables: Type I groups

g	Group	v	Variable
(1)	Demographic characteristics	1	Fraction population over 65
		2	Fraction population under 15
(2)	Economic system	3	Capitalism
		4	Socialism
(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 × log of GDP per capita
(4)	Relative government size	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

Notes: Variables that are not in the Sala-i-Martin *et al.* (2004) data set, but taken from Sala-i-Martin (1997) are indicated by a star (*). Group (8) is called 'Democracy' following Barro (1999).

TABLE 2

Grouping of variables: Type II groups

g	Group	v	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 resources	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

Notes: The representative variable of a group is indicated by a star (*).

TABLE 3

HWALS and WALS estimates: Focus variables

Variable	WALS-F8	HWALS-F8	
		Uniform prior	Data-dep. prior
Education		−0.0013 (0.0046)	0.0051 (0.0034)
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 × logGDP*			
Health		0.0073 (0.0058)	0.0062 (0.0059)
19 Life expectancy	0.0144 (0.0271)		
20 Malaria prevalence	−0.0045 (0.0094)		
Initial state		−0.0045 (0.0064)	−0.0084 (0.0057)
23 GDP in 1960 (log)	−0.0073 (0.0168)		
24 Size of economy	0.0006 (0.0186)		
Tropics effect		−0.0030 (0.0034)	−0.0041 (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)	−0.0030 (0.0028)	−0.0022 (0.0026)
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)	−0.0015 (0.0019)	−0.0022 (0.0018)
Price distortion			
70 Investment price	−0.0047 (0.0076)	−0.0041 (0.0017)	−0.0046 (0.0015)

Notes: A star (*) indicates that the variable is not in the Magnus *et al.* (2010) data set, so that no estimate for WALS-F8 is provided.

TABLE 4

HWALS and WALS estimates: Auxiliary variables

Variable	WALS-F8	HWALS-F8	
		Uniform prior	Data-dep. prior
Demographic characteristics		0.0027 (0.0048)	0.0026 (0.0044)
1 Frac. pop. over 65	−0.0011 (0.0204)		
2 Frac. pop. under 15	−0.0003 (0.0324)		
Economic system		−0.0010 (0.0016)	−0.0007 (0.0015)
3 Capitalism	0.0018 (0.0056)		
4 Socialism	−0.0000 (0.0067)		
Relative government size		−0.0003 (0.0021)	0.0005 (0.0020)
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)	0.0004 (0.0019)
21 Average inflation	0.0042 (0.0179)		
22 Square of inflation	−0.0064 (0.0200)		
Democracy		0.0025 (0.0027)	0.0015 (0.0024)
25 Political rights	0.0047 (0.0102)		
26 Civil liberties	0.0002 (0.0075)		
Scale effect		0.0028 (0.0028)	0.0018 (0.0024)
27 Land area	0.0063 (0.0157)		
28 Population	0.0005 (0.0086)		
Trade policy indices		0.0009 (0.0024)	0.0009 (0.0026)
29 Outward orientation	−0.0008 (0.0055)		
30 Years open	−0.0032 (0.0104)		
War		0.0001 (0.0016)	−0.0003 (0.0015)
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)	0.0006 (0.0029)	−0.0003 (0.0025)
48 Primary exports	−0.0026 (0.0104)		
Terms of trade			
49 Terms of trade ranking	0.0028 (0.0084)	0.0004 (0.0027)	0.0002 (0.0024)
50 Terms of trade growth	0.0026 (0.0058)	0.0035 (0.0024)	0.0021 (0.0022)
Regional effect			
51 East Asian dummy	0.0087 (0.0108)	0.0058 (0.0028)	0.0046 (0.0025)
52 African dummy	0.0017 (0.0117)	−0.0031 (0.0036)	−0.0020 (0.0032)
53 European dummy	0.0198 (0.0247)	0.0015 (0.0045)	0.0009 (0.0040)
54 Latin American dummy	0.0125 (0.0258)	−0.0014 (0.0046)	−0.0002 (0.0042)
55 Colony dummy	−0.0023 (0.0155)	−0.0038 (0.0035)	−0.0040 (0.0031)
56 British colony	−0.0003 (0.0071)	0.0028 (0.0027)	0.0022 (0.0026)
57 Spanish colony	−0.0015 (0.0152)	0.0012 (0.0033)	0.0007 (0.0031)

TABLE 4

HWALS and WALS estimates: Auxiliary variables, continued

Natural resources			
58 Hydrocarbon deposits	0.0015 (0.0053)	0.0001 (0.0019)	0.0005 (0.0017)
59 Frac. GDP in mining	−0.0016 (0.0072)	−0.0013 (0.0019)	−0.0012 (0.0017)
60 Oil country dummy	−0.0020 (0.0052)	−0.0018 (0.0023)	−0.0004 (0.0021)
Population			
61 Population density coastal	0.0019 (0.0172)	0.0007 (0.0030)	0.0026 (0.0025)
62 Interior density	−0.0025 (0.0070)	−0.0010 (0.0017)	−0.0008 (0.0015)
63 Fraction pop. in tropics	0.0003 (0.0092)	0.0015 (0.0032)	0.0009 (0.0028)
64 Population density	−0.0032 (0.0060)	−0.0015 (0.0021)	−0.0009 (0.0018)
65 Population growth rate	0.0073 (0.0232)	0.0014 (0.0054)	0.0003 (0.0047)
66 Fertility	0.0007 (0.0224)	−0.0030 (0.0063)	−0.0006 (0.0052)
Geography (excl. tropics effect)			
67 Frac. land area near water	0.0018 (0.0118)	0.0016 (0.0032)	0.0001 (0.0030)
68 Landlocked country dummy	0.0027 (0.0040)	0.0002 (0.0018)	−0.0003 (0.0016)
69 Air distance to big cities	0.0009 (0.0102)	0.0010 (0.0025)	−0.0001 (0.0023)
Real exchange rate			
71 Real exchange rate dist.	−0.0031 (0.0107)	−0.0021 (0.0020)	−0.0019 (0.0019)
Defense			
72 Defense spending share	−0.0145 (0.0599)	−0.0004 (0.0017)	−0.0007 (0.0016)
Political instability			
73 Revolutions and coups	0.0043 (0.0064)	−0.0006 (0.0018)	−0.0003 (0.0017)
Independence			
74 Timing of independence	−0.0001 (0.0110)	0.0008 (0.0025)	0.0010 (0.0023)

Notes: A star (*) indicates that the variable is not in the Magnus *et al.* (2010) data set, so that no estimate for WALS-F8 is provided.