

COMPARISONS OF MODEL AVERAGING TECHNIQUES: ASSESSING GROWTH DETERMINANTS

SHAHRAM M. AMINI AND CHRISTOPHER F. PARMETER

ABSTRACT. This paper replicates three important studies on growth theory uncertainty that employed Bayesian model averaging tools. We compare these results with estimates obtained using recently developed frequentist and alternative Bayesian model averaging techniques. Overall, we successfully replicate all three studies using freely available software in the statistical environment R, provide an easily implementable algorithm to operationalize the frequentist model averaging methods and find that the sign and magnitude of these new estimates are reasonably close to those produced via traditional Bayesian methods.

1. INTRODUCTION

The intrigue of uncovering a singular process that dictates economic growth has driven myriad economists to postulate growth theories explicating a link between a set of ‘determinants’ and a measure of economic growth. The propagation of theories on economic growth has simultaneously put forth an empirical conundrum known as “theory openendedness” (Brock & Durlauf 2001). Theory openendedness suggests that while numerous theories may indeed explain cross country growth, no one theory evinces a clear link with economic growth in the face of alternative growth determinants (both within and across growth theories). The lack of empirical robustness of growth models has sharpened the need for viable model selection and averaging techniques to parse through the vast amounts of data being used to empirically test competing growth models. Such techniques allow growth empiricists to tenably focus on the (set of) variables which produce robust relationships with economic growth.

For example, on a review of the literature considering alternative growth theories Durlauf, Johnson & Temple (2005) list over 140 different determinants which have been used in attempts to explain economic growth. The sheer volume of alternative regressors coupled with the limited number of countries presents quite a difficult econometric environment for growth empiricists.¹ To combat the array of potential growth determinants a common approach has been to employ model averaging econometric tools. Specifically, growth empiricists have used advances in both

Date: September 21, 2011.

Key words and phrases. Model Averaging, Mallow’s Criterion, Economic Growth, Enumeration, Model Sampling. Shahram M. Amini, Department of Economics, Virginia Polytechnic Institute and State University, email: shahram@vt.edu. Christopher F. Parmeter, Corresponding Author, Department of Economics, University of Miami, e-mail: cparmeter@bus.miami.edu. The authors would like to thank, but not implicate, Badi Baltagi, two anonymous referees and Klaus Moeltner for helpful advice and insightful criticism during the writing of this paper.

¹Not to mention additional econometric issues such as endogeneity, functional form specification, measurement error, etc.

computational power and speed to implement a variety of Bayesian model averaging (BMA) tools to uncover the effect that a given variable has on economic growth along with the posterior probability that a specific variable resides in the final ‘model’ explaining growth. Empirically, BMA methods have proved quite successful at delineating robust growth determinants and theories (see Durlauf, Kourtellis & Tan 2008).

However, for those more accustomed to operating within the frequentist paradigm, the appeal of BMA tools is diminished. Until recently, there existed a dearth of available frequentist model averaging (FMA) tools to directly compete with BMA. Hansen (2007) developed an easily implementable model averaging tool for linear regression known as Mallows model averaging (MMA) that has very desirable asymptotic optimality properties. These methods encapsulate model averaging within the frequentist econometric arena and are comparable to BMA tools. Thus, we seek to employ MMA to see how our results differ from similar analyses conducted using BMA within the growth theory uncertainty setting. Furthermore, we also consider a modified Bayesian estimator developed by Magnus, Powell & Prüfer (2010) which uses Bayesian insights to average over a set of frequentist estimators. This estimator is termed weighted average least squares (WALS).

Our focus will have several specific aims. First, we replicate the key results of three previous studies using BMA to evaluate various aspects of growth theory uncertainty: Fernandez, Ley & Steel (2001), Masanjala & Papageorgiou (2008) and Doppelhofer & Weeks (2009). This replication is important given the availability of different software to conduct BMA. Specifically, we focus on replicability within the computing environment R (R Development Core Team 2010). Within R there exist at least three alternative packages to conduct BMA, the **BMA** (Raftery, Hoeting, Volinsky, Painter & Yeung 2010), **BMS** (Zeugner & Feldkircher 2009), and **BAS** (Clyde 2010) packages. Replicability of empirical studies using BMA tools is important since the search algorithms deployed can be computationally involved and it is important to ensure that the final results presented are inherent of the underlying relationship and not that of the search mechanism.

We find that model averaging using the **BMS** package exactly replicates the results of Fernandez et al. (2001) and Doppelhofer & Weeks (2009) as well as the working paper version of Masanjala & Papageorgiou (2008). The **BMA** package is able to exactly replicate the results of Masanjala & Papageorgiou (2008) but neither of the other two studies, while the **BAS** package cannot replicate the results of any of the three studies.

Second, we directly compare the BMA posterior means and standard deviations of our three studies of interest with the model averaging methods (MMA and WALS) described above. We find that our final weighted estimated parameter vectors using both MMA and WALS are nearly identical to the posterior means obtained using BMA. However, for each of the three examples we encounter differences between the new methods’ standard errors and the standard deviation of the posterior distribution of each parameter stemming from the original BMA analysis. The difference in standard errors in the MMA setting most likely owns to the fact that a formal asymptotic theory does not yet exist for the MMA estimator and the method to construct standard errors are conservative. Third, we show that the model enumeration property of the WALS estimator (linear in

the number of covariates as opposed to exponential) also translates to the MMA estimator. Given that no appropriate sampling mechanism from the model space exists for the MMA estimator, this enumeration property makes implementation of the MMA estimator trivial for the number of covariates typically encountered in econometric exercises.

The remainder of our paper is as follows. Section 2 discusses the implementation of model averaging estimators across the menu of available methods. Section 3 briefly describes the data used for our replications. Section 4 presents our results for the BMA, WALs and MMA estimators and discusses differences and similarities across the estimates while Section 5 concludes.

2. MODEL AVERAGING TECHNIQUES

2.1. Bayesian Model Averaging. Given the prevalence of descriptive overviews on BMA (Raftery 1995, Hoeting, Madigan, Raftery & Volinsky 1999), we only briefly describe its implementation here. A BMA analysis begins by considering a set of possible models, say $M = \{M_1, M_2, \dots, M_K\}$, where M represents the model space the averaging will be conducted over. Once this space has been set the posterior distribution of the parameters of interest from a linear model (say), β , given the data, D , is

$$(1) \quad Pr(\beta|D) = \sum_{j=1}^K Pr(\beta|M_j, D)Pr(M_j|D),$$

which is simply the weighted average of the posterior distributions under each model. Using Bayes Theorem, one can rewrite the posterior probability of model M_k as follows

$$(2) \quad Pr(M_k|D) = Pr(D|M_k) \frac{Pr(M_k)}{Pr(D)} = Pr(D|M_k) \frac{Pr(M_k)}{\sum_{j=1}^K Pr(D|M_j)Pr(M_j)}$$

$$(3) \quad Pr(D|M_k) = \int Pr(D|\beta^{(k)}, M_k)Pr(\beta^{(k)}|M_k)d\beta^{(k)}$$

where $\beta^{(k)}$ is the vector of parameters from model M_k , $Pr(\beta^{(k)}|M_k)$ is a prior probability distribution assigned to the parameters of model M_k , and $Pr(M_k)$ is the prior probability that M_k is the true model. As mentioned above, in order to make BMA applicable, we need to choose prior distributions for the parameters of the model and then, based on the data, calculate posterior probabilities attached to different models and finally, find the parameter distributions by averaging the posterior distributions.

One should notice that, as stated in Brock, Durlauf & West (2003), the set of prior probabilities must be informative with respect to the likelihood, meaning priors should be relatively high where the likelihoods are large otherwise the choice of priors will have a substantial effect on the posteriors. Several studies have attempted to discern the effect that the choice of prior probabilities has on posterior probabilities, which are commonly used to gauge the ‘robustness’ of a variable. Ley & Steel (2009) provide a complete review on the effect of prior assumptions on the outcome of BMA analyses. Additionally, Eicher, Papageorgiou & Raftery (2011) address the issue of which default

prior to use in applied settings when information on the prior is lacking or small relative to the information in the data. They use cross-validated predictive performance and find support for the unit information prior (UIP) for the integrated likelihood coupled with a uniform prior over the model space. The key implication of their work is that the UIP is the safest and most robust choice to use as a benchmark. However, it is not clear that a specific prior will always emerge as the best prior to use across all econometric settings.

2.2. Weighted Average Least Squares. Magnus et al. (2010) recently introduced the WALS estimator based off of the Equivalence Theorem of Mean Squared Error (MSE), discussed in Magnus & Durbin (1999) and Danilov & Magnus (2004). The WALS estimator, which is a Bayesian combination of frequentist estimators, takes the required prior probabilities from the Laplace distribution and generates bounded risk, unlike standard BMA estimators (which typically deploy normal priors). To implement the WALS estimator, assume a linear regression model

$$(4) \quad y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon, \quad \varepsilon \sim i.i.d N(0, \sigma^2)$$

where y is an $n \times 1$ vector of the dependent variable, (β_1, β_2) is the vector of model parameters, X_1 ($n \times k_1$), $X_1 \neq \emptyset$, contains variables that will always be included in the model (such as an intercept), X_2 ($n \times k_2$) includes auxiliary variables which may be included in the model and $X = (X_1 : X_2)$, the matrix of regressors, has full column-rank.²

Initially we orthogonalize the columns of X_2 such that $Q'X_2'H_1X_2Q = \Lambda$, where $H_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$, Λ is a diagonal matrix, and Q is an orthogonal $k_2 \times k_2$ matrix. Consequently,

$$(5) \quad X_2\beta_2 = X_2Q\Lambda^{-1/2}\Lambda^{1/2}Q'\beta_2 = \tilde{X}_2\beta_2^*$$

This suggests that we can replace $X_2\beta_2$ in equation (4) with $\tilde{X}_2\beta_2^*$, estimate (β_1, β_2^*) rather than (β_1, β_2) , and then recover β_2 from $\beta_2 = Q\Lambda^{-1/2}\beta_2^*$. Given that any combination of the columns of X_2 can be used to evaluate (4), there are a total of $K = 2^{k_2}$ models to be evaluated. For model j , $k_{2j} \leq k_2$ columns of X_2 are selected. The estimators of β_1 and β_2 under model j , after some simple algebra, can be written as

$$\hat{\beta}_1^{(j)} = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2W_jX_2'H_1y, \quad \hat{\beta}_2^{(j)} = W_jX_2'H_1y$$

where W_j is a diagonal $k_2 \times k_2$ matrix with k_{2j} ones and $k_2 - k_{2j}$ zeros on the diagonal, such that the m th diagonal element is zero if the m th element of $\beta_2^{(j)}$ is zero. The WALS estimator of β_1 is simply a weighted average of the $\hat{\beta}_1^{(j)}$ s

$$\hat{\beta}_{1,WALS} = \sum_{j=1}^{2^{k_2}} \lambda_j \hat{\beta}_1^{(j)} = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2W X_2'H_1y$$

²The error terms are assumed to be distributed identically and independently while a more general assumption on the error term for the WALS estimator is discussed in Magnus, Wan & Zhang (2011).

where $W = \sum_j \lambda_j W_j$. The Equivalence Theorem states that

$$E(\hat{\beta}_{1,WALS}) = \beta_1 - (X_1' X_1)^{-1} X_1' X_2 E(W \hat{\beta}_2 - \beta_2), \quad \text{where } \hat{\beta}_2 = X_2' H_1 y$$

This means that $\hat{\beta}_{1,WALS}$ is a “good” estimator of β_1 if and only if $W \hat{\beta}_2$ is a “good” estimator of β_2 . $W = \text{diag}(\omega_1, \omega_2, \omega_3, \dots, \omega_{k_2})$ is a diagonal matrix and the components of $\hat{\beta}_2$ are independent because of the model space orthogonalization. Therefore, if we choose ω_j to be a function of $\hat{\beta}_{2j}$ (a scalar) only, then the components of $W \hat{\beta}_2$ will be independent. Namely, our 2^{k_2} -dimensional problem reduces to k_2 one-dimensional problems. The importance of this result is that the WALS estimator, unlike the BMA estimator, can fully enumerate the model space even for large k_2 since it is *linear* in the number of covariates under inspection. A standard BMA implementation is exponential in the number of covariates and even for a moderate number of covariates MCMC methods are used to sample from the model space since full enumeration is computationally expensive (if not impossible).

Using the fact that $\hat{\beta}_{2j} | \beta_{2j} \sim N(\beta_{2j}, \sigma^2)$, the Laplace estimator of β_{2j} is defined as

$$\tilde{\beta}_{2j} = E(\beta_{2j} | \hat{\beta}_{2j}) = \frac{1 + \delta(\hat{\beta}_{2j})}{2} (\hat{\beta}_{2j} - c) + \frac{1 - \delta(\hat{\beta}_{2j})}{2} (\hat{\beta}_{2j} + c)$$

where

$$\delta(\hat{\beta}_{2j}) = \frac{e^{-c\hat{\beta}_{2j}} \Phi(\hat{\beta}_{2j} - c) - e^{c\hat{\beta}_{2j}} \Phi(-\hat{\beta}_{2j} - c)}{e^{-c\hat{\beta}_{2j}} \Phi(\hat{\beta}_{2j} - c) + e^{c\hat{\beta}_{2j}} \Phi(-\hat{\beta}_{2j} - c)}$$

$\Phi(\cdot)$ represents the CDF of the standard normal distribution, and c is a positive hyperparameter that usually is set to be $c = \log(2)$ for the Laplace distribution. Define $\tilde{\beta}_{2,Laplace} = (\tilde{\beta}_{21}, \tilde{\beta}_{22}, \tilde{\beta}_{23}, \dots, \tilde{\beta}_{2k_2})$. Then the WALS estimators for β_1 and β_2 are

$$(6) \quad \hat{\beta}_{2,WALS} = \sigma Q \Lambda^{-1/2} \tilde{\beta}_{2,Laplace}$$

$$(7) \quad \hat{\beta}_{1,WALS} = (X_1' X_1)^{-1} X_1' (y - X_2 \hat{\beta}_{2,WALS})$$

See Magnus et al. (2010, Section 3.5) for further details on implementation of the WALS estimator.³

2.3. Frequentist Model Averaging. An alternative method to deal with model uncertainty is FMA. Unlike model selection approaches where the researcher has to select one model, according to a criterion, from the set of acceptable models $M = \{M_1, M_2, \dots, M_K\}$, and unlike BMA, which eventually computes the probability of being in the final model for all available variables, an FMA estimator assigns weights, based upon a specific criterion, to each model, M_j , $j = 1, 2, \dots, K$, and then takes the weighted average of all the parameter vectors to obtain the model average estimator. Mathematically, the estimator of the parameters of interest, β , is

$$(8) \quad \hat{\beta} = \sum_{j=1}^K \omega_j \hat{\beta}^{(j)}$$

³See also Magnus et al. (2011) for a formal application.

where ω_j , $j = 1, 2, \dots, K$ are weights normalized such that $\sum_j \omega_j = 1$, and assigned to the corresponding models, and $\hat{\beta}^{(j)}$ is the estimate of β under model j . Buckland, Burnham & Augustin (1997) suggest calculating the weight vector based on an exponential AIC criterion while Hansen (2007) suggests using a Mallows's criterion for the selection of the weights. The use of the Mallows's criterion is desirable since both Hansen (2007) and Wang, Zhang & Zou (2009) show that asymptotically the model average estimator in (8) achieves the lowest possible squared error in the class of model average estimators.⁴

To implement the MMA approach laid out in Hansen (2007) consider a standard linear regression model

$$(9) \quad y = X\beta + \varepsilon = \sum_{i=1}^{\infty} \beta_i x_i + \varepsilon, \quad \varepsilon \sim i.i.d D(0, \sigma^2)$$

where y is the $n \times 1$ vector of dependent observations and β is the vector of model parameters. Let X_{k_j} be the corresponding submatrix of X containing the first k_j columns of X .⁵ Index $j = 1, 2, 3, \dots, K$ indicates the sequence of models. We estimate the ℓ th model, $y = X_{k_\ell} \beta^{(\ell)} + \varepsilon_\ell$, by least squares and denote the vector of estimated coefficients under model ℓ as $\hat{\beta}^{(\ell)}$ and the vector of estimated residuals as $\hat{\varepsilon}_\ell$. This procedure is repeated for all K models and the final estimator of β is presented in equation (8), where ω_j , $j = 1, 2, \dots, K$ are normalized weights calculated numerically by minimizing the Mallows criterion,

$$(10) \quad C_p(\Omega) = \Omega' \hat{\varepsilon}' \hat{\varepsilon} \Omega + 2\sigma^2 L' \Omega$$

with $\Omega_{(K \times 1)} = (\omega_1, \omega_2, \dots, \omega_K)'$ representing the vector of weights, $\hat{\varepsilon}_{(n \times K)} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_K)$ is the matrix of all residual vectors, and $L_{(K \times 1)} = (k_1, k_2, \dots, k_K)'$ is the vector of the number of regressors used for each of the K models.⁶ This is, in fact, a classic linear-quadratic programming problem. Mathematically, the optimization problem is

$$(11) \quad \begin{aligned} \min_{\omega_j} C_p(\Omega) &= \Omega' \hat{\varepsilon}' \hat{\varepsilon} \Omega + 2\sigma^2 L' \Omega \\ \text{s.t.} \quad \sum_{j=1}^K \omega_j &= 1, \quad 0 \leq \omega_j \leq 1, \quad \forall j = 1, 2, \dots, K. \end{aligned}$$

A solution for this minimization problem can be found in any number of programming languages.⁷ One can replace the unknown σ^2 with the estimator $\hat{\sigma}_K^2 = (n - K)^{-1} \hat{\varepsilon}'_K \hat{\varepsilon}_K$. Hansen (2007, Theorem 2) shows that $\hat{\sigma}_K^2$ is a consistent estimator for σ^2 provided $K/n \rightarrow 0$ and $K \rightarrow \infty$ as the sample size increases. Moreover, replacing σ^2 with an estimator does not affect the asymptotic optimality properties of the weight vector selected via the Mallows criterion.

⁴See Li (1987) for a similar proof in the model selection context.

⁵Since we assume $E(\varepsilon|X) = 0$ the first column of X can be thought of as a vector of ones representing the intercept.

⁶A thorough review of MMA is provided in Wang et al. (2009).

⁷We use the `quadprog` package (Turlach & Weingessel 2007) available in R.

Unfortunately, no formal asymptotic theory exists for the MMA estimator to obtain the variance-covariance matrix needed to construct standard errors. We therefore employ Buckland et al.'s (1997) suggestion to calculate the variance of our MMA estimator based on the bootstrap. Their conservative approach finds the maximum possible value for variances that incorporate a component representing misspecification bias. Buckland et al. (1997) and Schomaker, Wan & Heumann (2010) propose estimating the variance of an FMA estimator as

$$(12) \quad \text{var}(\hat{\beta}) = \sum_{j=1}^K \omega_j \sqrt{\text{var}(\hat{\beta}^{(j)} | \text{bias}_j) + \text{bias}_j^2}$$

that accounts for the additional uncertainty due to model selection. In this formula $\text{bias}_j = \hat{\beta}^{(j)} - \hat{\beta}$ is the misspecification bias that arises in estimating β under model j . We use the full model, including all covariates, as the benchmark to calculate the misspecification bias. That is, we use OLS to estimate the vector of parameters, $\beta^{(j)}$, under model j , and also the variances of the estimated parameters. Then we subtract the vector of estimated parameters under model j from the full model to construct the bias for model j . As Hjort & Claeskens (2003) mentioned, this formula gives reasonably accurate standard errors that include additional uncertainty due to model selection.

2.3.1. Enumerating the Model Space. To implement the MMA estimator two key issues remain. First, Hansen (2007, Theorem 1) requires an explicit ordering of the regressors. Fortunately, Wan, Zhang & Zou (2010) established the optimality of MMA without requiring the explicit ordering needed in Hansen's (2007) proofs.⁸ Second, ignoring the ordering, the question of how best to select variables and combine them for all the models to be weighted within the Mallows's criterion is paramount. Here, we use the pioneering insight of Magnus et al. (2010) to reduce the total size of the model space from $K = 2^k - 1$ to $K = k$ where k is the largest number of covariates that one has. This result is obtained by showing that the key results of Hansen do not change when we have an orthogonalization of the entire set of covariates.

For the moment we will assume that σ^2 is known. Use the notation X_m to denote the matrix of covariates used to construct the OLS estimates for the m th model. The OLS estimator for the m th model is $\hat{\beta}^{(m)} = (X_m' X_m)^{-1} X_m' y$. The matrix $P_m = X_m (X_m' X_m)^{-1} X_m'$ will figure prominently in the following discussion. What we want is to have X_m such that

$$Q' X_m' X_m Q = \Lambda,$$

where Q is an orthogonal matrix and Λ is a diagonal matrix. We then have $\Lambda^{-1/2} Q' X_m' X_m Q \Lambda^{-1/2} = I$. Letting $\tilde{X}_m = X_m Q \Lambda^{-1/2}$ we have $\tilde{X}_m' \tilde{X}_m = I$. We note that Lemma 1 in Hansen (2007) still

⁸In results not reported here for our replication studies, we do indeed obtain approximately identical coefficient estimates when we consider different orderings of the regressors within our MMA framework. Furthermore, see Wan et al. (2010, Example 2) which shows how considering a set of nested models satisfies the requisite conditions for the non-ordering to hold, provided the model space grows slow enough.

holds because

$$\tilde{P}_m = \tilde{X}_m(\tilde{X}'_m\tilde{X}_m)^{-1}\tilde{X}'_m = \tilde{X}_m\tilde{X}'_m = X_mQ\Lambda^{-1}Q'X'_m = X_m(X'_mX_m)^{-1}X'_m = P_m.$$

Also, it is trivial to recover $\beta^{(m)}$ from $\beta^{*(m)}$ (the coefficient vector from the orthogonalization) since

$$\tilde{X}_m\beta^{*(m)} = \tilde{P}_m y = P_m y = X_m\beta^{(m)}.$$

This yields $\beta^{*(m)} = \Lambda^{1/2}Q'\beta^{(m)}$ so once we have estimates of $\beta^{*(m)}$ we can recover estimates of the parameters of interest as $\beta^{(m)} = Q\Lambda^{-1/2}\beta^{*(m)}$.

The importance of the orthogonalization is that the residuals, $\hat{\varepsilon}_m = H_m y = (I - P_m)y$ are identical across the K models that we estimate for the model averaging. The independence amongst the covariates means that when a new variable is added to the model, the information it contains is not corrupted by covariances with the other variables, implying that no additional combinations of the variables are needed to exhaust the model space. In effect, what the orthogonalization gives us is the ability to select a single variable and continue to add to the model space *linearly*. Also, given that the variables are orthogonal the order in which this is done is immaterial. The simplest way to orthogonalize the vector space is to use the singular value decomposition or Gram-Schmidt orthogonalization.

Once the orthogonalization has been undertaken the MMA weights can be found by simply estimating a set of models which increase by a single variable each time.⁹ For example, suppose that there were 5 regressors which one wanted to average over. The total number of models to average over would be $2^5 - 1 = 31$. However, if we orthogonalize the regressors then we have only 5 models to average over $\{1, (1, 2), (1, 2, 3), (1, 2, 3, 4), (1, 2, 3, 4, 5)\}$. This may seem trivial for a small number of regressors, but it quickly becomes computationally relevant. With just 20 regressors, not uncommon in applied work, we already have more than one million elements of the model space. It is clear that both MMA and WALs can be implemented across a large number of covariates without concern that enough models are visited, as in BMA investigations.

3. DATA

Our basis for comparison of BMA, WALs and FMA techniques are three recent studies of growth theory uncertainty. We first replicate the results of Fernandez et al. (2001), Masanjala & Papageorgiou (2008) and Doppelhofer & Weeks (2009) using freely available software that conducts BMA. Second, we compare the BMA posterior means and standard errors to those obtained using both the MMA and WALs estimators described above for the set of variables deemed important in each paper. All three datasets are publicly available on the Journal of Applied Econometrics online data archive. Due to the fact that these datasets are well documented we briefly discuss them here and direct the reader to more detailed treatments.

Our first study, Fernandez et al. (2001) (FLS hereafter) uses a cross section of 72 countries along with 41 potential growth determinants for the period 1960 to 1992. This dataset is taken from

⁹Given the independence of the regressors the order in which this is done is irrelevant.

the larger dataset used by Sala-i-Martin (1997) for his study on robust determinants of growth. Our second dataset is a subset of the FLS data used by Masanjala & Papageorgiou (2008) (MP hereafter). MP’s focus was on growth theory uncertainty and if the determinants of growth differed for African countries relative to the rest of the world. They used the 37 sub-Saharan African countries within the FLS dataset along with 24 of the 41 variables. See the appendix in MP for the countries and variables used in this study. Our last dataset comes from Doppelhofer & Weeks (2009) (DW hereafter) who used the data in Sala-i-Martin, Doppelhofer & Miller (2004) to study jointness of growth determinants. This data set comprises 88 countries and 67 candidate variables in a cross section for the period 1960 to 1996. All 67 of the variables used can be found in data appendix B in DW.

4. RESULTS

To implement the BMA methods in the three papers of interest we first consider the similarity of the results from the 3 available packages currently available in R (**BMA**, **BAS**, and **BMS**). Amini & Parmeter (2011) detail the features of each of these three packages and advocate on behalf of the **BMS** package. Table 1 presents the posterior inclusion probabilities (PIP), posterior means and standard deviations from each of the packages using the Fernandez et al. (2001) study as a benchmark. For succinctness, in the remainder of the paper we only present results for variables with PIPs greater than 0.5 from the published studies. As is apparent, only the **BMS** package is successful at matching the reported PIPs and posterior mean/standard deviation while the **BAS** package comes close to reproducing the posterior means/standard deviations but the PIPs display significant differences (compare the results for Rule of Law and Years Open Economy). The **BMA** package also produces similar posterior means but cannot achieve the same accuracy of the PIPs. Given these results, coupled with the recommendation of Amini & Parmeter (2011) we present all remaining BMA results using the **BMS** package.¹⁰

In passing we note that the results in Table 1 are not due to the specific study under comparison. For all three of the studies considered here the **BAS** package cannot replicate the full set of estimates (PIPs, posterior mean/standard deviations) while the **BMA** package can only reproduce the results of the Masanjala & Papageorgiou (2008) study, primarily because this study used the **BMA** package for estimation.¹¹ Our intuition for this breakdown between the **BMA**, **BMS** and **BAS** packages is that the **BMA** package does not use a Monte Carlo Markov Chain (MCMC) algorithm to search over the model space¹² and the available options within the packages does not allow the priors to be identical across the three packages; see Amini & Parmeter (2011) for more details on the options available within each of these packages.

¹⁰PIPs and posterior mean and standard deviations stemming from the additional packages are available upon request.

¹¹A working paper version of Masanjala & Papageorgiou (2008) possesses different results than the published version which, coincidentally, the **BMA** package cannot fully replicate while the **BMS** package can.

¹²It evaluates the most visited models based upon a Bayesian Information Criterion difference and a hierarchical OLS t -test that is different from the methods used in both the **BMS** and **BAS** packages.

TABLE 1. Performance of the BMS, BAS, and BMA Packages for the Fernandez et al. (2001) study.

Variable:	Published Results	BMS Package		BAS Package		BMA Package	
	PIP (i)	PIP (ii)	Post Mean & SD (iii)	PIP (iv)	Post Mean & SD (v)	PIP (vi)	Post Mean & SD (vii)
GDP 1960	1.00	1.00	-0.0160 (0.0031)	1.00	-0.0168 (0.0026)	1.00	-0.0173 (0.0026)
Fraction Confucian	0.995	0.990	0.0563 (0.0147)	1.00	0.0667 (0.0115)	1.00	0.0590 (0.0109)
Life Expectancy	0.946	0.941	0.0008 (0.0003)	0.999	0.0008 (0.0002)	1.00	0.0008 (0.0002)
Equipment Investment	0.942	0.939	0.1594 (0.0685)	0.985	0.1373 (0.0446)	0.972	0.0993 (0.0440)
Sub-Saharan Dummy	0.757	0.751	-0.0118 (0.0084)	0.996	-0.0211 (0.0051)	1.00	-0.0184 (0.0053)
Fraction Muslim	0.656	0.634	0.0084 (0.0077)	0.430	0.0049 (0.0068)	0.519	0.0055 (0.0061)
Rule of Law	0.516	0.510	0.0074 (0.0082)	0.942	0.0110 (0.0048)	1.00	0.0126 (0.0043)
Years Open Economy	0.502	0.500	0.0069 (0.0079)	0.049	0.0001 (0.0014)	0.055	0.0003 (0.0018)

Notes: This table presents the results of applying Bayesian model averaging to the data used in Fernandez et al. (2001) using the BMS, BAS, and BMA packages. The dependent variable is the growth rate from 1960-1996 across 72 countries. Standard errors are in parenthesis. All models estimated contain an intercept. Column (i) shows the Bayesian posterior inclusion probabilities that Fernandez et al. (2001) have reported. Columns (ii), (iv), and (vi) report the PIPs using BMS, BAS, and BMA packages, respectively. Columns (iii), (v), and (vii) indicate BMA posterior estimates of the model coefficients and standard errors using the three packages.

For the following analyses we construct BMA, WALs, and MMA estimators using the three datasets, FLS, MP, and DW, and compare the findings. Implementing the BMA estimator requires specifying the set of model priors and a search algorithm. We use the uniform distribution for model priors and a MCMC algorithm to search over the model space to construct our BMA estimator. Our implementation of the WALs estimator begins by orthogonalizing the entire dataset and then using the moments of the Laplace distribution explained in section 2.2 to compute the appropriate weights using $c = \log(2)$ as the hyperparameter. We use a similar methodology to implement the MMA estimator. We add one variable linearly at a time to the set of covariates, orthogonalize the vector space constructed by those variables, and then compute the weights by minimizing the Mallows's criterion provided in (11). The total number of models visited by both WALs and MMA is k , the number of covariates in each dataset, given our discussions in sections 2.2 and 2.3.1. Regardless of model averaging estimator (BMA, WALs, MMA) all models estimated contain an intercept.

4.1. A comparison using Fernandez et al. (2001). Table 2 presents the BMA, WALs and MMA estimates for the Fernandez et al. (2001) study for the variables with PIPs above 50%. Column (i) presents the PIPs, calculated via the BMS package. The explanatory variables are sorted from the highest PIP to the lowest. Column (ii) reports the estimated posterior means and their standard errors. Column (iii) reports the model average estimates and standard deviations using the WALs estimator of Magnus et al. (2010) while column (iv) reports the corresponding model average estimates using the MMA estimator of Hansen (2007) using the orthogonalization described earlier.

Comparing across the three set of estimates we highlight that both the sign and magnitude of all three sets of estimates are identical. However, the three estimators deliver different point estimates and standard deviations for each of the variables. While not a strict relationship, the MMA point estimates appear to be marginally larger in absolute value for all seven variables relative to the WALS estimates (excluding Fraction Muslim). A strict relationship between the three sets of estimated standard errors does not appear to exist.

One important highlight from the comparison of the three estimation approaches here is that the results are relatively ‘close’ and part of this in term stems from the fact that all three frameworks use a uniform prior on the model space. The ability to place differing probability mass on different models throughout the model space is a strength of the Bayesian approach. While the researcher can estimate, or feed in specific models to the WALS and MMA setups, the full generality of the model prior is an advantage of using BMA. This is an important issue that, while beyond the scope of this paper, certainly requires further study within the FMA estimation paradigm.¹³

TABLE 2. BMA, WALS, and MMA estimation for Fernandez et al. (2001) study.

Method:	PIP-BMA	BMA	WALS	MMA
	(i)	(ii)	(iii)	(iv)
Variable:				
GDP 1960	1.00	-0.0160 (0.0031)	-0.0139 (0.0029)	-0.0169 (0.0032)
Fraction Confucian	0.990	0.0563 (0.0147)	0.0663 (0.0141)	0.0747 (0.0150)
Life Expectancy	0.941	0.0008 (0.0003)	0.0007 (0.0002)	0.0008 (0.0002)
Equipment Investment	0.939	0.1594 (0.0685)	0.1041 (0.0378)	0.1272 (0.0445)
Sub-Saharan Dummy	0.751	-0.0118 (0.0084)	-0.0157 (0.0061)	-0.0201 (0.0067)
Fraction Muslim	0.634	0.0084 (0.0077)	0.0108 (0.0079)	0.0106 (0.0086)
Rule of Law	0.510	0.0074 (0.0082)	0.0089 (0.0058)	0.0116 (0.0061)

Notes: This table presents the Bayesian model averaging (BMA), Weighted average least squares (WALS), and Mallows model averaging (MMA) techniques applied to the data used in Fernandez et al. (2001). The dependent variable is the growth rate from 1960-1996 across 72 countries. Standard errors are in parenthesis. All models estimated contain an intercept. Column (i) reports the Bayesian posterior inclusion probabilities (PIPs), calculated via the *BMS* package. Column (ii) reports the Bayesian posterior means and their standard errors. Column (iii) reports the estimation of model coefficients and standard deviations using the WALS approach discussed in Magnus et al. (2010). Column (iv) reports the corresponding model averaging estimates from the MMA estimator of Hansen (2007) using the orthogonalization of model space. The BMA analysis was set up with a uniform prior on the model space.

¹³We reexamined the PIPs and posterior means from our BMA results using two alternative sets of priors. First, we used “fixed” priors corresponding to fixed common prior inclusion probabilities for each regressor, as described in Sala-i-Martin et al. (2004). In this case the PIPs and posterior means are almost identical to the current case employing uniform priors. Second, we used “random” priors that trigger the ‘random theta’ prior by Ley & Steel (2009), who suggest a binomial-beta hyperprior for the prior inclusion probability. In this case the PIPs and posterior means are significantly different from the other two settings. The reason is that using random model priors is recommended if strong prior information on model size exists. Overall, it seems that, as stated in the introduction, results stemming from a BMA analysis are sensitive to the choice of model priors. Additionally, the fact that the posterior means differ substantially also suggests that examining MMA against a nonuniform prior BMA approach may not necessarily represent a valid comparison.

4.2. A comparison using Masanjala & Papageorgiou (2008). Masanjala & Papageorgiou (2008) study the sources of growth in Africa. This was an important study because not only did it address model uncertainty, but it also focused on parameter heterogeneity. Their results find key determinants of growth in Africa that are substantially different from those in the rest of the world. They conclude that mining, primary exports, and primary school enrollment are the main drivers behind sub-Saharan African growth. The published version of the paper produces results using the BMA package and as such replication of this paper is trivial.¹⁴

Table 3 reports the same set of comparison results as Table 2 for Masanjala & Papageorgiou's (2008) study, using those variables with estimated PIPs higher than 50%. Column (i) presents the PIPs. The explanatory variables are sorted from the highest PIP to the lowest. Column (ii) reports the estimated posterior means and their standard errors. Column (iii) reports the model average estimates and standard deviations using WALS while column (iv) reports the corresponding model average estimates using the MMA estimator with the variable orthogonalization described earlier. We mention again that replication of Masanjala & Papageorgiou (2008) is trivial given that they used the BMA package.

A key difference that arises between the BMA point estimates and those of WALS and MMA are the variables Years Open Economy and Mining. First, Years Open Economy has a positive effect using BMA while it has a negative effect using either WALS or MMA. Second, Mining has a positive effect for all three model averaging methods, yet, the magnitude of the WALS point estimate is almost three times larger than that of BMA while the MMA results are more than three times larger than the posterior means from BMA. Even more important, both WALS and MMA suggest that Mining is statistically relevant, as judged by the t -ratio while the BMA results suggest statistical insignificance. Contrary to what we witnessed in Table 2, here the three competing methods yield different insights.

4.3. A comparison using Doppelhofer & Weeks (2009). Doppelhofer & Weeks (2009) recently focused on model averaging exercises where jointness was considered. Jointness of explanatory variables is based on the joint posterior distribution of variables over the model space as opposed to marginal posterior distributions, which were the focus of the previous two studies. Moreover, jointness is an important concept to capture because it impacts statistical inference and can be informative for economic policy. Doppelhofer & Weeks (2009, Appendix B) provides the BMA PIPs, posterior means and standard errors. Table 4 presents the estimates for the three different model averaging methods. The BMA results are exactly replicated from the Doppelhofer & Weeks (2009) study. Similar to what we witnessed from Table 2, the results from the BMA, WALS and the MMA methods all produce coefficient estimates which are similar in sign and magnitude.

¹⁴However, the BMA package was unable to replicate the results of the working paper version of MP whereas the BMS package was able to obtain identical estimates. Moreover, the variables that are relevant are different between these two versions of the paper suggesting that differences in model priors and search algorithms necessary for estimation can produce incongruous insights.

TABLE 3. BMA, WALS and MMA estimation for Masanjala & Papageorgiou (2008) study.

Method:	PIP-BMA	BMA	WALS	MMA
	(i)	(ii)	(iii)	(iv)
Variable:				
GDP 1960	0.824	-0.0106 (0.0060)	-0.0162 (0.0041)	-0.0195 (0.0041)
Years Open Economy	0.732	0.0309 (0.0215)	-0.0264 (0.0275)	-0.0367 (0.0317)
Mining	0.670	0.0398 (0.0328)	0.0928 (0.0305)	0.1242 (0.0348)
Primary Exports	0.538	-0.0208 (0.0221)	-0.0160 (0.0165)	-0.0258 (0.0187)
Primary School Enrollment	0.514	0.0005 (0.0006)	-0.0001 (0.0005)	-0.0001 (0.0005)

Notes: This table presents the Bayesian model averaging (BMA), Weighted average least squares (WALS), and Mallows model averaging (MMA) techniques applied to the data used in Masanjala & Papageorgiou (2008). The dependent variable is the growth rate from 1960-1996 across 37 sub-Saharan African countries. Standard errors are in parenthesis. All models estimated contain an intercept. Column (i) reports the Bayesian posterior inclusion probabilities (PIPs), calculated via the BMS package. Column (ii) reports the Bayesian posterior means and their standard errors. Column (iii) reports the estimation of model coefficients and standard deviations using the WALS approach discussed in Magnus et al. (2010). Column (iv) reports the corresponding model averaging estimates from the MMA estimator of Hansen (2007) using the orthogonalization of model space. The BMA analysis was set up with a uniform prior on the model space.

A consistent theme arising from all three studies is that the standard errors from MMA are slightly larger than those from WALS and both methods have standard errors roughly double than those found using BMA. One possible explanation for this is that the construction of the variances of the coefficient estimates for the MMA estimator are conservative given that it is not based on a formal asymptotic representation. An alternative explanation, at least for the DW study is that for the implementation of the MCMC model search, the BMS package only needs to visit 29,000 models out of 2^{68} models to achieve convergence. Compare this with the FLS dataset, where the BMS package needs to visit 52,000 models out of 2^{42} to achieve convergence, which in percentage terms is more than 100 million times larger than the percentage of potential models visited in the DW dataset using the same MCMC model search algorithm. In this sense the BMA results for the DW study are highly accurate, using models with very high likelihood. Moreover, this fact coupled with the full enumeration of the model space by WALS and MMA means that many poor models could have been used which adds to the uncertainty of the average coefficient estimates, resulting in larger standard errors.

Ignoring the root cause of the discrepancies in the standard errors across the three methods, even with the conservative nature of the MMA standard errors, the significance of the weighted estimates for MMA across the three examples is roughly identical to those stemming from WALS and are broadly consistent with the BMA results. A formal derivation of the limiting distribution of the MMA estimator marks an interesting avenue for future research.

5. CONCLUSIONS

This paper set out to replicate several important studies focused on model uncertainty in the modeling of economic growth. Current research on this influential area within growth empirics has

TABLE 4. BMA, WALs, MMA estimation for Doppelhofer & Weeks (2009) study.

Method:	PIP-BMA	BMA	WALS	MMA
	(i)	(ii)	(iii)	(iv)
Variable:				
East Asian Dummy	0.848	0.0182 (0.0094)	0.0116 (0.0155)	0.0132 (0.0193)
Primary Schooling Enrollment	0.798	0.0215 (0.0128)	0.0204 (0.0163)	0.0304 (0.0205)
Investment Price	0.790	-0.00006 (0.00003)	-0.00008 (0.00004)	-0.0001 (0.00005)
Initial GDP 1960	0.742	-0.0062 (0.0045)	-0.0053 (0.0072)	-0.0081 (0.0091)
Fraction of Tropical Area	0.591	-0.0085 (0.0078)	-0.0090 (0.0154)	-0.0144 (0.0193)

Notes: This table presents the Bayesian model averaging (BMA), Weighted average least squares (WALS), and Mallows model averaging (MMA) techniques applied to the data used in Doppelhofer & Weeks (2009). The dependent variable is the growth rate from 1960-1996 across 88 countries. Standard errors are in parenthesis. All models estimated contain an intercept. Column (i) reports the Bayesian posterior inclusion probabilities (PIPs), calculated via the *BMS* package. Column (ii) reports the Bayesian posterior means and their standard errors. Column (iii) reports the estimation of model coefficients and standard deviations using the WALS approach discussed in Magnus et al. (2010). Column (iv) reports the corresponding model averaging estimates from the MMA estimator of Hansen (2007) using the orthogonalization of model space. The BMA analysis was set up with a uniform prior on the model space.

almost solely relied on BMA techniques to date. Aside from our goal of replication we also compared the BMA results to similar estimates obtained using alternative Bayesian methods (WALS) and recently introduced FMA techniques (MMA). These newer methods may be appealing to some because they allow for full enumeration of the model space which avoids having to invoke MCMC search algorithms to sample from the model space. Additionally, the FMA methods do not require the specification of prior distributions to make the averaging estimator operational. For our FMA estimator we employed the Mallows's criterion to select the weights as it has been shown to deliver asymptotically optimal weights. Using freely available software to implement the Bayesian methods we were able to exactly reproduce the results of all three studies. Additionally, we found that the sign and magnitude of both the WALS and the MMA coefficients were roughly identical to those obtained via BMA.

We are encouraged by the similarity of these three methods as well as our ability to replicate the results of these computationally intensive estimation methods. However, several important caveats are noted. The primary focus of our replication/comparison was based on the weighted coefficient estimates and their standard errors. However, traditional BMA analyses also focus attention on the PIP of a variable as well as measuring jointness between pairs of covariates. In the current setup, neither WALS or MMA provide this information. This suggests an extant relative strength of BMA over its competitors. Future work to derive similar metrics from WALS and MMA are required before having the ability to favor one method over another categorically.

Beyond our comparison and replication of three important studies into the determinants of economic growth, our implementation of the MMA estimator required use of orthogonalization of the covariate space to fully enumerate the model space, making the analyses computationally feasible. This orthogonalization is the same insight that makes WALS appealing in an applied

context. Orthogonalization makes it possible to easily implement model averaging estimators in the presence of a large number of covariates and should prove useful in an array of economic contexts.

REFERENCES

- Amini, S. & Parmeter, C. F. (2011), Bayesian Model Averaging in R. University of Miami, Department of Economics Working Paper WP2011-9.
- Brock, W. A., Durlauf, S. N. & West, K. D. (2003), ‘Policy Evaluation in Uncertain Economic Environments’, *Brookings Papers on Economic Activity* (1), 235–323.
- Brock, W. & Durlauf, S. (2001), ‘Growth Empirics and Reality’, *The World Bank Economic Review* **15**(2), 229–272.
- Buckland, S. T., Burnham, K. P. & Augustin, N. H. (1997), ‘Model selection: an integral part of inference’, *Biometrics* **53**(2), 603–618.
- Clyde, M. (2010), *BAS: Bayesian Adaptive Sampling for Bayesian Model Averaging*. R package version 0.92.
URL: <http://CRAN.R-project.org/package=BAS>
- Danilov, D. & Magnus, J. (2004), ‘On the harm that ignoring pretesting can cause’, *Journal of Econometrics* **122**(1), 27–46.
- Doppelhofer, G. & Weeks, M. (2009), ‘Jointness of growth determinants’, *Journal of Applied Econometrics* **24**(2), 209–244.
- Durlauf, S. N., Johnson, P. & Temple, J. (2005), Growth econometrics, in P. Aghion & S. N. Durlauf, eds, ‘Handbook of Economic Growth’, Vol. 1A, North Holland, Amsterdam, pp. 555–677.
- Durlauf, S. N., Kourtellos, A. & Tan, C.-M. (2008), ‘Are any growth theories robust?’, *Economic Journal* **118**, 392–346.
- Eicher, T. S., Papageorgiou, C. & Raftery, A. E. (2011), ‘Default priors and predictive performance in Bayesian model averaging with application to growth determinants’, *Journal of Applied Econometrics* **26**, 30–55.
- Fernandez, C., Ley, E. & Steel, M. (2001), ‘Model uncertainty in cross-country growth regressions’, *Journal of Applied Econometrics* **16**(5), 563–576.
- Hansen, B. (2007), ‘Least squares model averaging’, *Econometrica* **75**(4), 1175–1189.
- Hjort, N. & Claeskens, G. (2003), ‘Frequentist model average estimators’, *Journal of the American Statistical Association* **98**(464), 879–899.
- Hoeting, J. D., Madigan, D., Raftery, A. & Volinsky, T. (1999), ‘Bayesian model averaging: A Tutorial’, *Statistical Science* **14**, 382–417.
- Ley, E. & Steel, M. (2009), ‘On the effect of prior assumptions in Bayesian model averaging with applications to growth regression’, *Journal of Applied Econometrics* **24**(4), 651–674.
- Li, K. (1987), ‘Asymptotic optimality for C_p , C_L , cross-validation and generalized cross-validation: discrete index set’, *The Annals of Statistics* **15**(3), 958–975.
- Magnus, J. & Durbin, J. (1999), ‘Estimation of regression coefficients of interest when other regression coefficients are of no interest’, *Econometrica* **67**(3), 639–643.
- Magnus, J. R., Powell, O. & Prüfer, P. (2010), ‘A comparison of two model averaging techniques with an application to growth empirics’, *Journal of Econometrics* **154**(2), 139–153.
- Magnus, J. R., Wan, A. T. K. & Zhang, X. (2011), ‘Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market’, *Computational Statistics and Data Analysis* **55**, 1331–1341.
- Masanjala, W. & Papageorgiou, C. (2008), ‘Rough and Lonely Road to Prosperity: A reexamination of the sources of growth in Africa using Bayesian Model Averaging’, *Journal of Applied Econometrics* **23**(5), 671–682.
- R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>
- Raftery, A. (1995), ‘Bayesian model selection in social research’, *Sociological Methodology* **25**, 111–163.

- Raftery, A., Hoeting, J., Volinsky, C., Painter, I. & Yeung, K. Y. (2010), *BMA: Bayesian Model Averaging*. R package version 3.13.
URL: <http://CRAN.R-project.org/package=BMA>
- Sala-i-Martin, X. (1997), 'I just ran two million regressions', *The American Economic Review* **87**(2), 178–183.
- Sala-i-Martin, X., Doppelhofer, G. & Miller, R. (2004), 'Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach', *American Economic Review* **94**(4), 813–835.
- Schomaker, M., Wan, A. T. K. & Heumann, C. (2010), 'Frequentist model averaging with missing observations', *Computational Statistics and Data Analysis* **54**, 3336–3347.
- Turlach, R. & Weingessel, A. (2007), *quadprog: Functions to solve Quadratic Programming Problems*. R package version 1.4-11.
- Wan, A. T. K., Zhang, X. & Zou, G. (2010), 'Least squares model averaging by Mallows criterion', *Journal of Econometrics* **156**, 277–283.
- Wang, H., Zhang, X. & Zou, G. (2009), 'Frequentist model averaging estimation: A review', *Journal of Systems Science and Complexity* **22**(4), 732–748.
- Zeugner, S. & Feldkircher, M. (2009), 'Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging', *IMF Working Papers* .